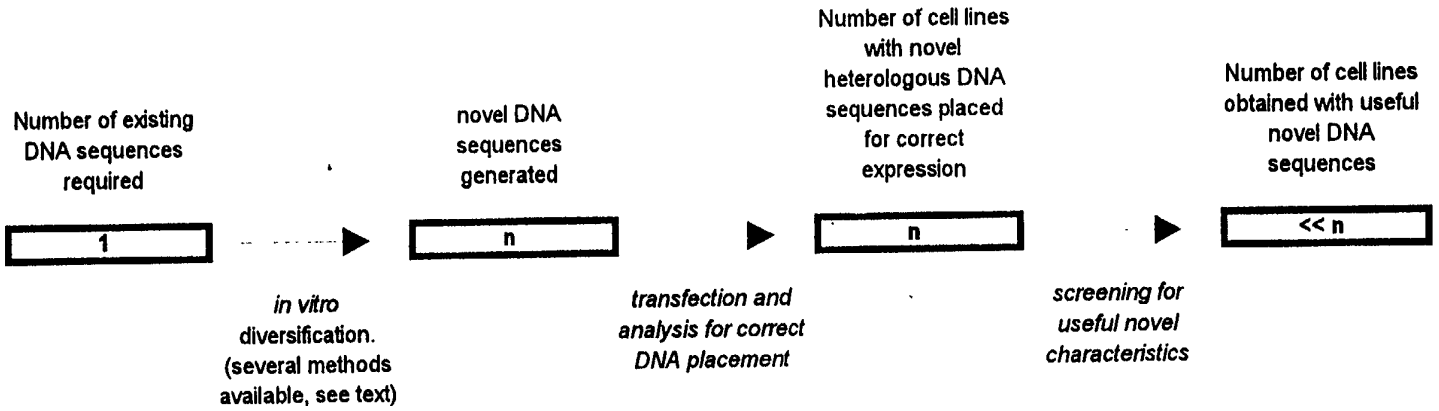
Figure 1 Methods for the diversification of DNA sequences and testing for superior variants

1A existing protocols: ***Number of transfections needed to generate 1024 new variants: 1024***



1B a protocol enabled by the present invention: ***Number of transfections needed to generate 1024 new variants: 2***

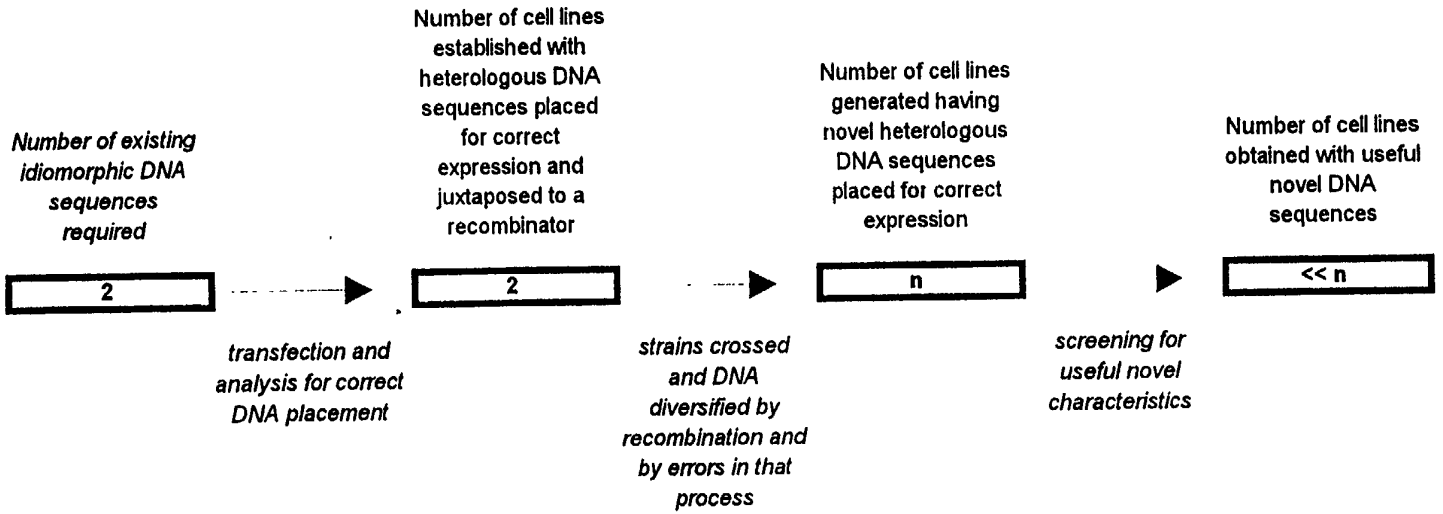


Figure 2 Methods for the diversification of DNA sequences coding subunits of heteropolymeric proteins and testing for superior variants.

The example given for immunoglobulins is for illustrative purposes only and is not intended to limit application of the present invention to this specific heteromeric protein. H = heavy chain genes, L = light chain genes

Figure 2A existing protocol: ***Number of transfections needed to generate 1024 new combinations: 2048***

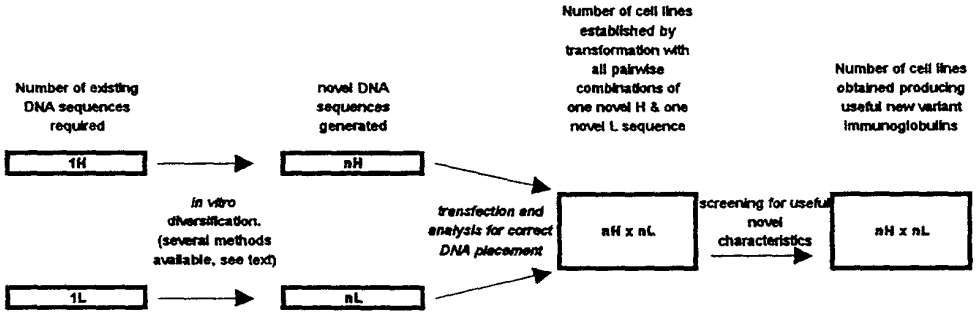


Figure 2B existing protocol using the heterokaryon technology of US Patent Serial No. 5,643,745
Number of transfections needed to generate 1024 new combinations: 64

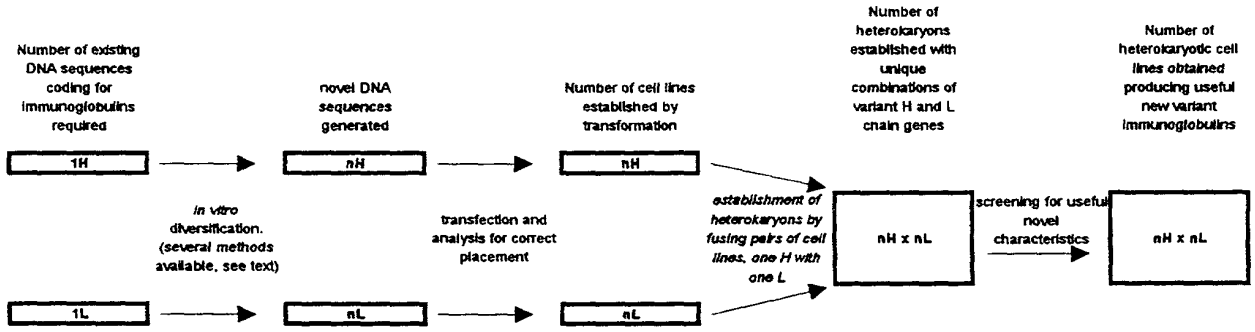


Figure 2C a protocol enabled by the present invention
Number of transfections needed to generate 1024 new combinations: 4

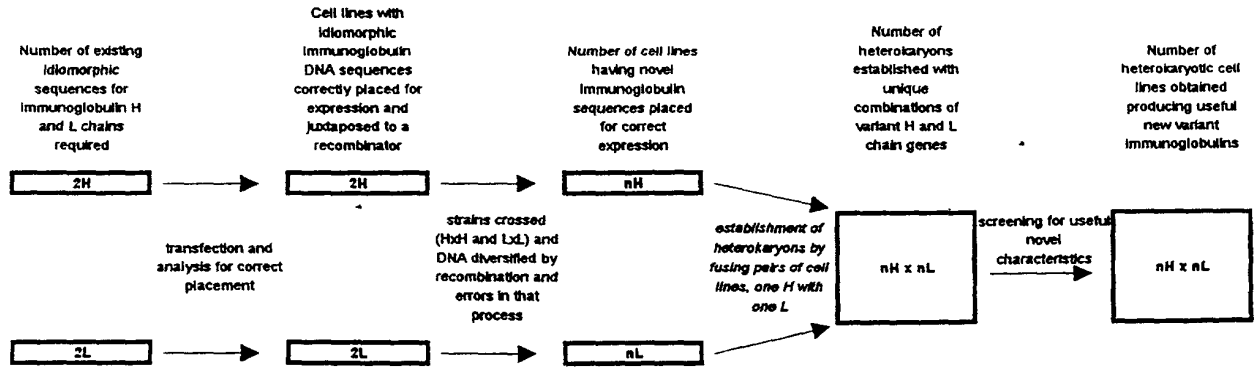


Figure 3 The modified double strand break repair model for meiotic recombination. After H Sun *et al* Cell **64:** 1155-1161, 1991

(a) A double strand break (DSB) is made in one DNA duplex. (b) A long 3' overhanging single strand tail is generated either side of the break by resection. (c) One 3' end invades a homologous duplex forming a D loop. (d) the D loop is enlarged by repair synthesis and anneals to the second 3' end. (e) Repair synthesis occurs at the second 3' end and two intermolecular junctions (Holliday junctions) are formed. Resolution of the junctions by cutting inner and outer strands can give rise to non-crossover (f) and crossover (g) chromosomes. If there are base mismatches in the heteroduplex regions (duplex molecules with thick and thin lines) there will be gene conversion. If mismatch repair does not occur there will be post meiotic segregation of new sequence combinations.

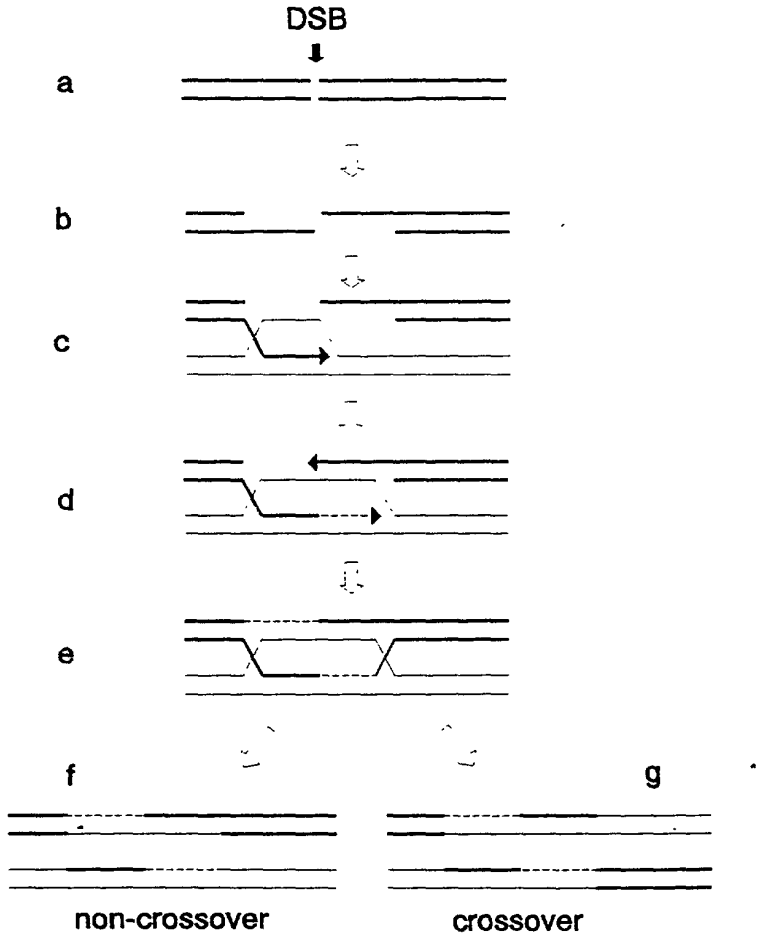
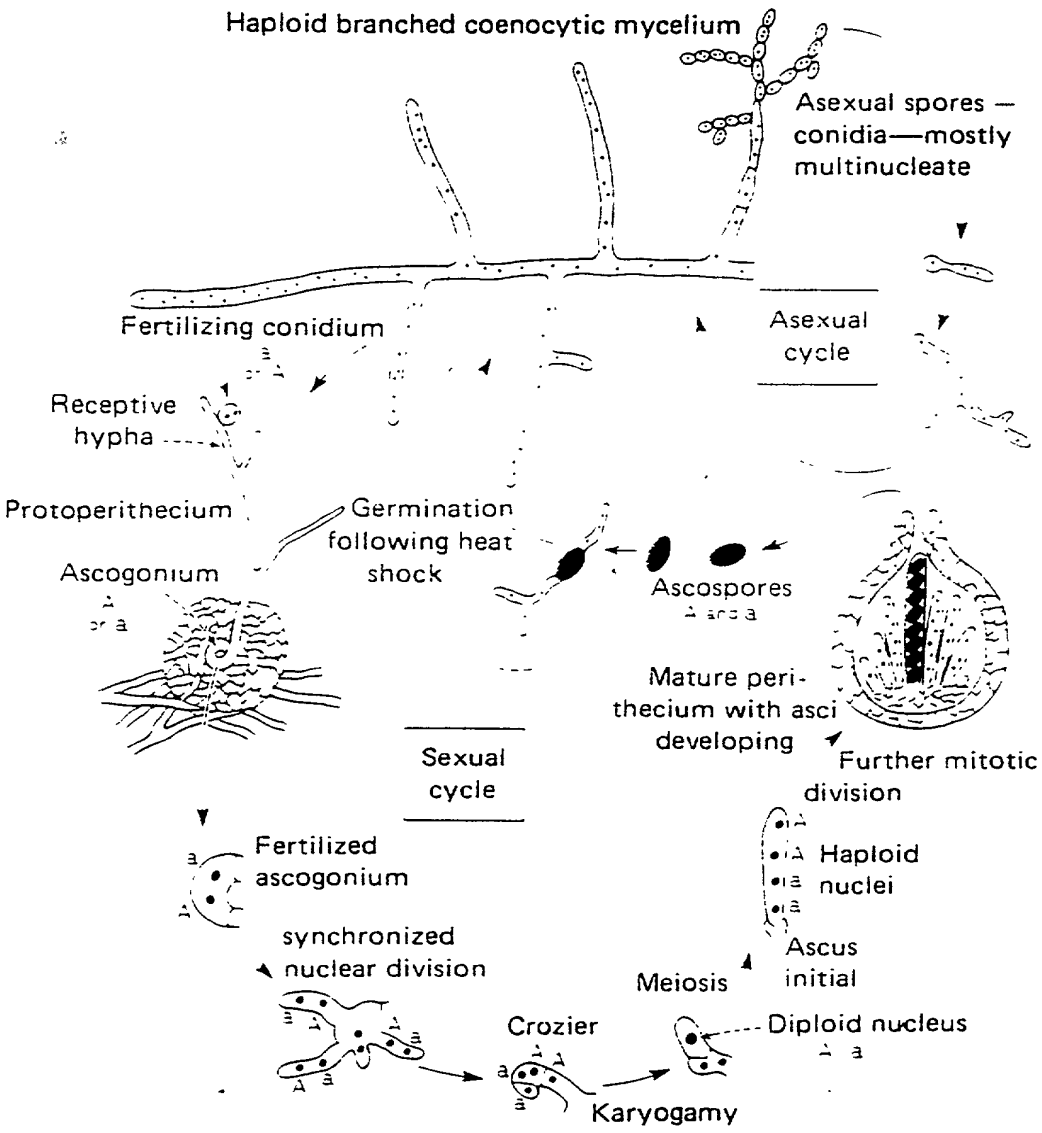
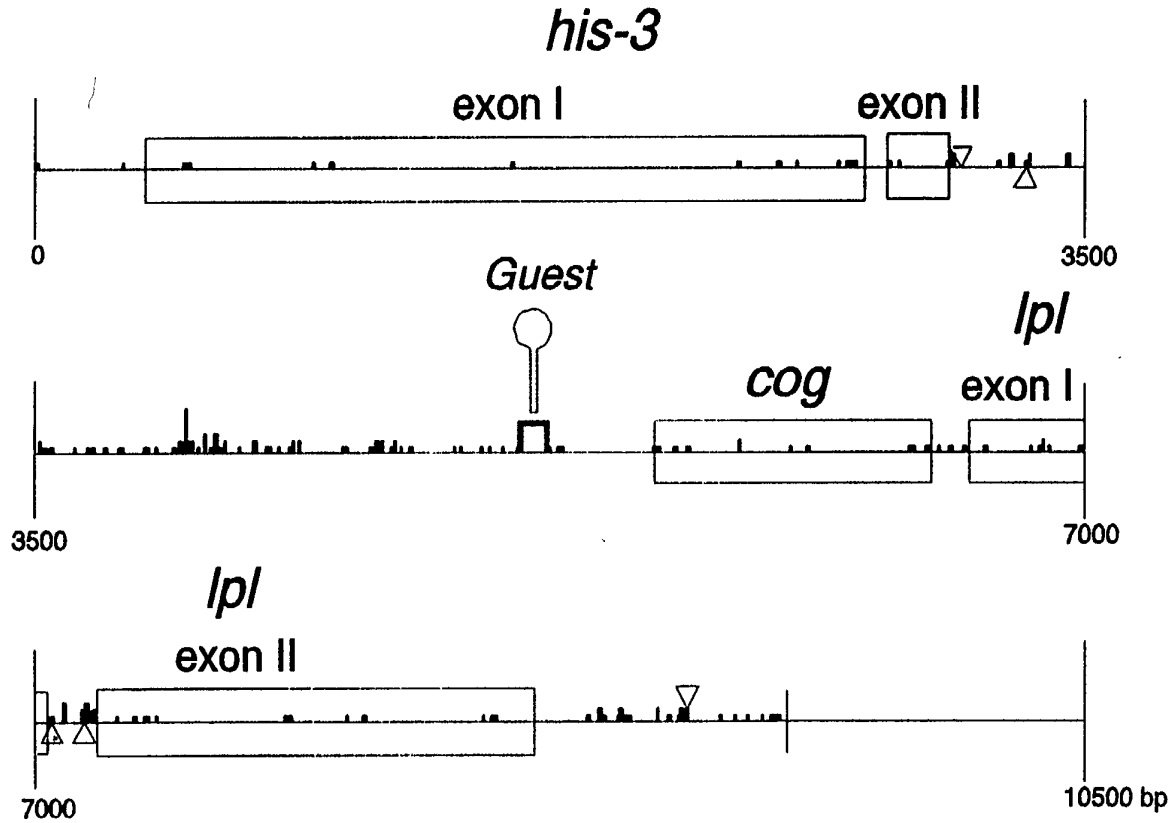


Figure 4 Life cycle of *Neurospora crassa* after JRS Fincham (Genetics, Wright 1983). Microconidia having one nucleus are not shown but can be generated as described in the text. Perithecia and protoperithecia are shown in section.



006337 3567560

Figure 5 Map of the *his-3, cog, lpl* region of Linkage Group I of *Neurospora crassa*. Vertical bars , triangles and hairpins show the location of sequence differences that distinguish the St Lawrence and Lindegren wild type strains. The corresponding full DNA sequences are given in figure 7 and figure 8. Vertical slashes indicate one to seven base substitutions per 10 base pairs. Triangles indicate short sequence insertions and the hairpin a 101 base pair inverted repeat transposon fragment present in St Lawrence.



006237 2956252.60

Figure 6. Discontinuity in the parental origin of DNA sequences in progeny from crosses between pairs of *his-3* alleles. In most cases this reflects discontinuity of conversion tracts, in some cases crossovers near the ends of conversion tracts.

Markers are specific DNA sequence differences that distinguish the parents. These were all E (Emerson wild type origin) or all L (Lindegren wild type origin) in the parental strains. Recombinants carry both E and L markers.

Marker position is given in base pairs from the first base of the first codon of the *his-3* gene. Each line of the table shows the parental origin of the markers inherited by one of the progeny.

marker	P	H	P1	K1201	K504	L3	R1	K26	K874	R4	C4	C5	C6	C6/7	C7	C8	C1	C2	C3	C9	D
location		~-600	-384	115	179	563	1232	1502	1717	2318	3436	3705	4000	4304	4667	4821	5232	5495	6153	6507	

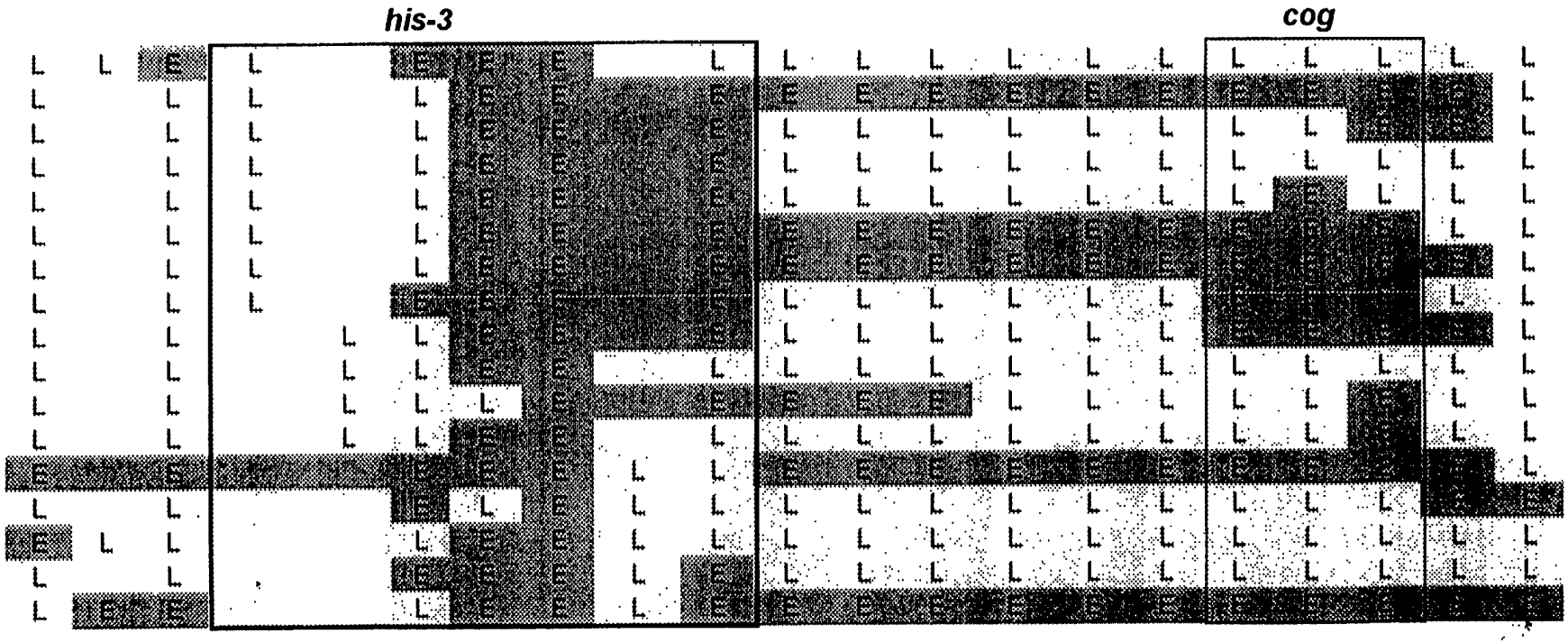


Figure 7 Nucleotide sequence of the *his-3 cog^{L} lpl* region of linkage group I in the Lindegren wild type strain of *Neurospora crassa*. This differs from that in the StLawrence strain in many positions, summarised in figure 5. The coordinates of relevant features are given in the text. This sequence contains the high frequency recombinator *cog^{L}* which is active providing the cross in which meiosis occurs is homozygous *rec-2*.

```
   1 GATCGCAACT GGAGATCACT CGCACCGTGC CGCAGAACAA GGGCGACGAG CCTCAGGGCA
  61 GTTTAGCCTG CCGTAACAGC ACAGACCATA GCTTATTTTC ACCTGGGCGG GCGGGCGACG
 121 GCGGCACTGA CATCGGCAAG GCGGCATCAA GCAACCCCTC TGTTGCTTGC CAGCTGCCGG
 181 CCAACGTCAG CGGTACAAGG AGAAATCTGG AAGGAAAGAC TTCTGGCACC GACAGGATGG
 241 CACGCGGGAA AAGTTCCCAA TGCATGAGAT GAGGGGCATT TGCATTGCCT CCCGTCACAC
 301 TGCCCGCGAA CCCCAACCCC ACCATAGCGT CTGTCGATAC ATGGAGCGCG AAGTCGAGAA
 361 ACCTGTAATT CCTGGTAACT TTCAGGTACA CAGTACGTAC TGATCCTGGT ATCAAACCTT
 421 GCCTGCCGAG TTTTCGACGG AAAGAGGTGT GAATTGTGAA AGAGTCATAC CAAATCACCC
 481 GATTTTCATA AAGCCCGAGT CTTTTCTGTA CATAAGCGAC ACTCGAAGCG GGCCTCATCT
 541 TCATAGCCTG ATAGCTTGTA ATACTCCATC CTCGTATCTC ACTTGACCTT GAGTTCAACC
 601 CCACGTCAGA CTTCACCCGA CACATCGACG GATTGGGGAA CAGCACAATA CCTGAAAAGC
 661 GAGAAAACCA AACAGAGGAA AACACCATGG AGACAACACT TCCCCTCCCC TTCCTCGTCG
 721 GTGTCAGTGT TCCTCCCGGA CTGAATGACA TCAAGGAGGG CCTCAGCCGG GAGGAAGTCT
 781 CGTGTCTTGG CTGCGTCTTC TTCGAGGTCA AGCCCAAGAC CCTTGAGAAA ATCGTGCGAT
 841 TCCTCAAGCG TCACAATGTC GAATTTGAGC CCTACTTCGA TGTAACAGCC CTCGAGTCTA
 901 TCGATGATAT TATCACTCTT CTGGACGCCG GCGCCCGCAA GGTGTTTGTC AAGACCGAGC
 961 AGTTGGCCGA CCTCTCCGCA TATGGCTCCC GCGTTGCCCC CATTGTCACT GGAAGCAGCG
1021 CTGCTTTGCT TTCCTCCGCC ACCGAGAGCG GCCTTTTGCT CTCCGGCTTC GATCAGACTG
1081 CCTCCGAGGC TGCACAGTTT CTGGAGGAGG CCAGAGACAA GAAAATTACC CCCTTCTTCA
1141 TCAAGCCCGT TCCTGGGGCC GATCTCGAAC AGTTCATCCA GGTCGCCGCC AAGGCTAACG
1201 CCATCCCCAT CCTGCCATCC ACTGGCTTGA CAACAAAGAA GGACGAGGCC GGAAAGCTTG
1261 CCATCTCCAC CATCCTCTCG AGCGTCTGGA AGTCTGACCG TCCCGATGGT CTGCTCCCCA
1321 CCGTTGTCGT TGATGAGCAC GACACTGCTC TGGGTCTGGT CTACAGCAGT GCCGAGAGTG
1381 TGAACGAGGC CCTCAGGACA CAGACTGGTG TCTATCAGAG CCGGAAGCGC GGTCTCTGGT
1441 ACAAGGGTGC TACTTCCGGA GACACTCAGG AGCTCGTCCG CATCTCGCTT GACTGCGATA
1501 ACGATGCTCT CAAGTTTGTC GTGAAGCAGA AGGGTCGTTT CTGCCACCTC GATCAGTCCG
1561 GCTGCTTTGG TCAGCTCAAA GGCCTTCCCA AGCTCGAGCA GACTTTGATT TCGAGGAAAC
1621 AGTCTGCCCC CGAGGGCTCC TACACTGCCC GTCTCTTCTC CGATGAGAAG CTAGTCCGGG
1681 CCAAGATCAT GGAGGAGGCT GAGGAGCTCT GCACCGCTCA GACCCCCCAG GAAATCGCCT
1741 TTGAGGCTGC CGATCTCTTC TACTTTGCTC TTACCAGGGC CGTTGCTGCC GGCGTTACTC
1801 TTGCCGATAT CGAAAGGAGC CTTGACGCCA AGAGCTGGAA GGTCAAGCGC AGGACTGGAG
1861 ATGCTAAGGG TAAGTGGGCT GAGAAGGAGG GCATCAAGCC TGCGGCGTCC GCTCCCGCTG
1921 CCACTTCGGC CCCTGTCACC AAGGAGGCCG CCCAGGAGAC CACCCCTGAG AAGATCACCA
1981 TGAGACGTTT CGACGCCTCC AAGGTCTCTA CCGAGGAGCT CGATGCTGCT CTCAAGCGTC
2041 CTGCGCAAAA GTCGTCCGAT GCCATCTACA AGATCATTGT CCCCATCATC GAGGACGTCC
2101 GCAAGAACGG CGACAAGGCT GTTCTGTCGT ACACTCACAA GTTCGAGAAG GCTACCTCTC
2161 TTACTAGCCC CGTCCTGAAG GCGCCCTTCC CCAAGGAGCT TATGCAGCTC CCTGAGGAGA
2221 CCATTGCTGC CATCGACGTG TCCTTCGAGA ACATCCGCAA GTTCCACGCC GCCCAGAAGG
2281 AGGAGAAGCC CCTCCAGGTC GAGACCATGC CCGGTGTTGT CTGCAGCCGT TTCTCTCGTC
2341 CCATCGAGGC CGTCGGCTGC TACATCCCCG GCGGTACCGC CGTTCTCCCC AGCACTGCCC
2401 TTATGCTGGG TGTTCCCGCC ATGGTCGCCG GCTGCAACAA GATTGTGTTC GCCTCTCCTC
2461 CCCGCGCCGA CGGAACCATC ACTCCCGAGA TTGTCCACGT CGCTCACAAG GTTGGGGCCG
2521 AGTCCATCGT GCTTGCCGGC GGTGCCCAGG CCGTAGCTGC CATGGCCTAC GGCACCGAGA
2581 GCATCACCAA GGTCGACAAG ATTCTCGGCC CCGGTAACCA GTTCGTCACT GCTGCCAAGA
2641 TGTTCGTCAG CAACGACACC AACGCTGCCG TTGGGATTGA CATGCCCGCT GGCCCGTCCG
2701 AGGTGCTGGT CATCGCTGAC AAGGACGCCA ACCCCGCGTT CGTTGCCTCG GATCTCCTGT
2761 CCCAGGCTGA GCACGGCGTT GACAGTCAGG TCATCCTGAT CGCTATTAAC CTCGACGAGG
2821 AGCATCTTCA GGCTATTGAG GACGAGGTTC ACCGTCAGGC TATGGAGCTT CCTCGCGTCC
2881 AGATTGTCCG TGGCTCCATC GCCCACTCGA TCACCGTGCA GGTCAAGACC GTCGAGGAGG
2941 CCATGGAGCT CAGCAACAAG TACGCTCCTG AGCACTTGAT CCTCCAGATC AAGGAGGCCG
3001 AGAAAGCTGT CGATCTTGTC ATGAACGCTG GTAGTGTCTT CATTGGCGCT TGGACTCCTG
3061 AGTCCGTTGG CGATTACTCT GCTGGTGTTA ACCACTCGCT GCGTAAGTTA CATATCATAA
3121 ATAGCCCCGC TTCACAGATT CTTCTGCTAA CGTCAAGACA CATAGCTACC TATGGTTTTG
3181 GCAAGCAGTA CTCTGGCGTC AATCTCGCCT CGTTCGTCAA GCACATTACC AGCTCCAACT
3241 TGACTGCCGA GGGTCTCAAA AACGTCGGCC AGGCTGTCAT GCAGTTGGCT AAGGTTGAGG
3301 AGCTCGAGGC TCACAGAAGG GCGGTCAGCA TCCGTCTTGA GCACATGAGC AAGAGCAACT
```

Figure 7 continued

```
3361 AGACGGAAAT TCTTTTTCGA AGTTGCAAAA AAAACAAGAA CAAAAGGATG TAGTGGGTTG
3421 ATGTATATCT GGGTCATTTT GGGCACATAG AGTAATGATA ACGAGTTTTG GACATTGTAC
3481 TGTTCTGTAC AGGCTGAAGA TCAGTACATG AATCTGTTGG TAAGTGTAGA GACCCAAACG
3541 TCCCTTGAGT TTTTCTCCCT GTTCCAGAGA GGTGCTCGTC CCTGGGTGTT TATTTTCATT
3601 ATTACATCAA CCTTTTATTT TATTTTATTT TTTATTTTAC TTTTTTTTCC TTTTTTTCAG
3661 ATCATGCGTA CATGAACGGG GGAAGCACAG ACGATCGAAA CGTGGATGTC ACAATGTCGC
3721 TGCAGTGATG CTGCATTGCA TGAAGCGCCC ATCTCAATAT ACTTGCAGTC TTGCGCGTTG
3781 CACGTGAACT TCCCAAACAA CCGAATAAAA GACGGCGAAA AATGAAGATA AAAAAAAACC
3841 ATAATAAAAA TCGGAGGGAG TGTGGGAAAT GGTTTCTTTT AGCATTTAGA CCCCATAGCC
3901 GTGCACGCCC GGGTACAGAC AGGTTCATCG ATGTTGACAT TGACTGGGAC ACCAGGTCTA
3961 TCTATTTCAT CTCCTGTCCT CTACCATACA TCGGGACATC GGACATCTCG CTGTACCCCC
4021 CACACCCACA AAGTCTTATA AAAGCGCCAC ACCCGAGGAG GTTCGGTCGG CCCCACGAAC
4081 TCCGTGCCTC CCTGCCTGTT TACAGGGACC GAACGCTGGA GAAGCTTAGT TTCCTGACAT
4141 CCGGCCTACC CGAGCAGGAA AAGGGACAGC TCATAGGCGA GGAGGGATTT GAAGATGGGG
4201 ACATTTTGGA TGATTCGAGA GGAGGAACTA GGTACTGTAT CATGATAGTT CGGGGCAGCA
4261 TCTTGGCTGG GACATTGTTA ATACCTCGAT ATGATGAAGT GGGAGGGAGT TTTTTCATGT
4321 CTTGCCCAAG TCCCACTAAT CTTTTTTTTT TTTTGTACCA ACACCCAAGA TTCGGAGAAT
4381 AGTGTAAGGA TTCGCATTCA CAAGTGGAAG TCTGAGGATC TTTTTATATC TTTGTCTTCC
4441 GCGGACTGTT AACGATCCTA CAGCGAGCGA GCGAGCGGTC GGATGCGCTG ATCTGATAGG
4501 TGCAATATAC GGCCGCTTTC TCCGGTCGTG TAGTGTAAGC TCTGTCGGCA TAGTAGTACA
4561 CTAAAAAAAC CCTTGCATTT CATGATCTGC TTGCTATTCA TTCCGAGTTA TTTCAGTGGT
4621 CACATTTCGA GATTCACAGC CATCCATCCA TATGGAAAAA TCCATTCCCA TGCTTCCTCC
4681 CCCCCACTAT GTATGTGACC ACACGCTGCT GTCAGAATGC CAACGGTCTC AGGTACCCTC
4741 GTCCGACTGT TTGGCATGGA GTTACATACA CTACTAGTGT AGCCCCGGGC CAAGCTACCC
4801 CGTCAAATCT ATACATATCT ATAATGGGTT TCAGGTGTTT CGTTCGCTGT CAATCAAGTT
4861 TGAAACATCA CTGGGGCCGT TGGACGGTGT ATTAGACCAT TGGCTCCCTC AGCTGGCGGC
4921 TGGGCGGTTG GGTCGGCAAT AACGGGACTG GACTTGAGAG GGACGAGGAG AGTCGGTTGG
4981 CTGCCTACAC TACACTACAA GCGTTCCCAC CTAACCGACG AGTCCCGTTT TCCATTTGTG
5041 TGCCTTAACC ATCATCTAGG GATGTCAGGG TTTGGCCGGA TCAGGGTATG TTTGGTTGAC
5101 TGTTGTCATG TCTGATTGGG TACATATCAT GGTAGGTGTC TCGAGAACAG TAGAGTACTC
5161 GGGCCTAGCG TTTGGATGAT TACGCGAGAT ATGAGTTGTA GGCCGCCATG CAGTTGCTTG
5221 CCCATAAGCA GAAGTTGCTT TGGGATATAT TTCTCGTCTT TCAAAGGTCA CGAGGTCCTG
5281 GGACGAGCGG CATCGCCATC CAAAGGGTTG AACATGAGAA ACCGGAATGG CCTTTGCGTT
5341 GAAATACAAA AAGTCAAGAA TAAAATCGCT TGAGGATAGG GACGTGGAAG CAAGCAAATA
5401 TGGTAAGGGA GGTACTGCTA TGTAGGTGCT CAGCAAACTG CCAATTTCTT GGCCCCCAAG
5461 CAGCAGTTTG CTGTCAGTGC TGCTCGTGTC AGCCTTGGTA GTGGAACCTA AACTGCTAAC
5521 ACAGCGCAAG TGCGCATGTA AAGATATTGT GGGAGGATCT GTATGGATGG ATGAGATTAC
5581 TGCTTGGTGT TGGTTGCGAG GCACTGCGGC TGTTAGGCTT TGCTGTGCCC CGTTCGACGA
5641 AGAAATACGC GGAACTATAA ATTGGATACC TAGACTTACT GCCTATGGGA GGTATCTACC
5701 GACGTAGCCG ACGGATTCTA GCAACATCCC GACTTTGCTT GTAGTGTACT ATGATAGCAG
5761 CACAGTGGGG TGTTGCTCCT TGTGAGCATG GGCTCTTTTT TTTTTTTTCC CCCTTCCCTA
5821 GGGCGTTGAC TGGACTTGCT CTATCGTTCC CAAGGTAGGT GCCCGTCATC GATTTTCCCA
5881 AGCCGTCTCC CGCCAGATTG TCGTCATAGT GTCATGATGA CCTCGGTCGC TGGGGCTGCG
5941 TGGTTACGGG GAGCTGGGAC CGCTAGGCCT CAGTGGTTGT GCCATTCAGC GTGGGTGTGT
6001 GGAGTAGCGG TAGAGGCGCT TGGAAGTTGT GCTAGCGGAA ACCCTGGAAT ATCTTGTACC
6061 CTTCGATTCC TTCTCGGGCT GCCCATGTGC TGAGGTGATG CCGGGGATCT GGCGCCAATC
6121 ATCCATTGAG GTTCCCGCAG CTTCCCGGTG CCGCGCGCGG GCGCAGTTGC TCACAGGACA
6181 CACCTAGACG CAGGGGCACA GGGGCACCGT TTGGTGTGCA ACTGGGTACC TGGTAGCTGT
6241 AGCAAGCACT CCACCGTCTG TGCAATCCCC CAATCCACGG CAGGAACTTA GCACCGCCGC
6301 GGCACCGAGT GAGCGAATCC ATCCGCATTG GATCCCAATT CTTGCCCTTG CCATCCTTCT
6361 TTCTTCCCAC TTGGCGCAAC CAACACTTCC CTTGGTCTGG GTACTCGTGT TGATCTTCAC
6421 TCTCTTTTTT TCTTGGGCGA CCGACTTTTT ATATCCGTCC TTGCTTCCCC CTGGCCGTTG
6481 TCGTTCTTTC TACAACTACC TTCCGTTCAT TATCCCCTTT CTTGGTTCGG TCGAGGACCC
6541 AAAAACAGAA CAATTCCGGC TCTTCCAGGT GGCTTGGGTG CGACTGTTTA GCTCTTGACC
6601 ACTAGCCGCT TACCTTCTCT TGATGTTTAT ATTTGGATAT CATTGAACTA CTCTTTCTTG
6661 AAACGGCAGA CGAACGGAAC AGTCCCTACG GTTTATTAGC GATATACGTT GTACTGATAT
6721 CCTGAGCAAG AAGAGGCAAA TTATCAATTA TGCATCTCCC ATCGTCGCTG CTCATCGCAG
6781 CTCCCTTGCT CGCCAATGTA TCGGCCGAAC CGATTAGGAT ACCCCAACGC GATGTTCTCC
6841 GTGGTATCAA CATCACAGCA ACTTGCCGTT CGAGCACTAC CGAATTCGCC CAGCGGTGGA
6901 TATGCCCCTG CCGTTGTAGA CTGTCCCAAG ACCAAGCCGA CGCTCCGGAA GGCCGTGGAT
6961 TTGTCGAACG AGGAGAAGAA CTGGTTGTCG ATCCGGAGGA AGAACACCAT CCAGCCCATG
7021 AGGGACCTAC TGAAGAGGGC CAACATCACT GGGTTCGATT CCGAAACTTT CATGAATGAG
7081 GCCGCCAACA ACGTCTCGCA ACTGCCCAAT GTCGCCATTG CCATTTCAGG AGGCGGCTAT
7141 CGTGCCCTCA TGAACGGCGC CGGCTTCGTT GCTGCTGCGG ATAACCGGAT TCAAAATACC
7201 ACGGGCGCAG GTGGTATTGG AGGCTTGTTG CAGTCCAGCA CATATTTGTA TGTAAAACCA
7261 TGCCTTCTTG TGGTTCTTCT TATCTCGTTT TCGAGTGTCA ACTGCGCCAG TTCGACGTTG
```

Figure 7 continued

```
7321 GGCGGCTGTG GACGACCTTG CTGGTGAACA TGTCTTGGAC TCCATGCCCC TTTTTTCGTT
7381 CCCTAAAATC CCAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAATTCGAG
7441 GACCGTGACT GTAAATTGCT AACGCAACTC TAGGGCCGGA CTTTCTGGTG GTGGCTGGCT
7501 TGTCGGCAGT TTGTTCTCCA ACAACTTCAG TAGCATTGAG ACCCTGCTGA GCGAGAACAA
7561 AGTCTGGGAC TTTGAGAACT CCATCTTTAA AGGACCCAAG GAGGCTGGCC TTAGTACTGT
7621 CAACCGTATC CAGTACTGGT CCGAAGTGGC AAAGGAAGTT GCGAAGAAGA AGGATGCTGG
7681 CTTCGAGACA AGTATAACAG ACTACTGGGG CCGAGCATTG AGTTACCAAC TGATCGGAGC
7741 CGATATGGGC GGCCCGGCTT ACACCTTCTC CAGCATTGCC CAGACCGACA ACTTCCAGAA
7801 GGCCGAAACG CCGTTCCCTA TTCTGGTAGC TGACGGCCGC GCGCCTGGAG ACACCATCAT
7861 CTCCCTCAAT GCTACCAACT ACGAGTTCAA CCCGTTCGAG ACGGGTAGCT GGGACCCGAC
7921 CGTCTATGGC TTTGCGCCGA CCAAGTACCT CGGCGCCAAC TTCAGCAACG GCGTGATCCC
7981 ATCGGGAGGC AAGTGCGTTG AGGGTCTCGA CCAAGCCGGC TTCGTCATGG GCACCAGCAG
8041 CACGCTCTTC AACCAGTTCC TTTTGGCCAA CATCTCCAGC TACGACGGTG TTGCCAGACG
8101 TGCTCATCGA GGCCGTGACT TCTGTCCTCA AGGAAATCGG CGCCAAGAGG ACGACGTCTC
8161 CCAAATCATC CCTAATCCGT TCCTGGACTG GAACAACCGG ACCAACCCCA ACGCCGACAC
8221 GCTCGAGCTC GACCTGGTCG ACGGCGGCGA AGATCTGCAG AATATTCCGC TCAACCCGCT
8281 CACCCAACCC GTGCGCGCCG TCGACGTCAT CTTCGCTGTC GACTCGTCCG CCGACGTGAC
8341 AAACTGGCCC AATGGCACCG CCCTGCGCGC CACCTACGAG CGCACTTTCG GCTCTATTTC
8401 CAACGGGACA CTCTTCCCCT CGATCCCCGA CGACTGGACG TTTATAAACC TAGGCCTCAA
8461 CAACCGCCCC TCTTTCTTCG GCTGCGATGT TAAGAACTTT ACCTTGAACG CCAACCAAAA
8521 GGTTCCCCCC TTAATCGTCT ATGTCCCCAA CGCGCCCTAT ACCGCGCTGA GCAACGTGTC
8581 CACCTTCGAT CCGTCATACA CGATGTCTCA GCGCAACGAC ATCATCGGCA ACGGATGGAA
8641 CTCAGCCACG CAGGGAAACG GCACGCTGGA TTCGGAGTGG CCCACTTGCG TCGCCTGCGC
8701 GGTTATCAGC AGGAGCTTAG ATCGGTTGGG CAGGCAGACG CCAGCCGCGT GCAAGACTTG
8761 CTTTGACAGG TATTGCTGGA ATGGCACAGT GAACTCCAAA GATACGGGGG TTTACATGCC
8821 TGAGTTCAAG ATTGCGGATG CGCATGCCCT GGACTCGGGT GCTGTTGCTA TCGGAAAGAT
8881 GGTGAATGTC TGGTCGTCGG TTGTGGTGGG AGTTGTGGCG GCTACTTTGT TGTTGTAGGG
8941 GTAGGGGAGA CGTGATGATA TTCCAGTCTG ATGAAGTTGA GACTGGACTG GAGATCGCCA
9001 AGGATGCGGA GGGAAAGGAA TGCGTGGTGT TAATGTCATG ATGGATGAAG AGTCATGGAT
9061 CATGGAACGA CGGGGCGGGG ATATTGGATG ATGGATATAC CACACTGCAT GCATGCTCTA
9121 TTGATAGTAT GCTTTGGCAT TTACGTTTAA CAATCAATTG CTCCATCCTG ATGTTCTATC
9181 TTTTTCGACA ATGGATTGAT ACTACTCCTG TTGCTTCGCT CTTGAGGTTG GAAGGACTTG
9241 AGGTTGGAAG GACTTGAGGT TGTTTGTTCT GAGGGAGGTT ATCGAAGTAT CATCTGTGCT
9301 GATGCCGATT GATAGACTGT CCTCTTCTTC GAGGCAACGA ACGGTCGGAT GAGCCTCTTT
9361 AATCATGATG CTCAGTGCCA CAAAAAGGCT CCAGCACAGC TGCCCACACC TTTCTTGCCT
9421 CGCCGTTCCT TCCTTTTTCT TTTCCCCTGT TTCCTTTCTT CCTTTCCATC TCATCCCGTA
9481 CCAGAGTGCC CACCGGGTAT ATATATTACC TCCTTGGCCG TTCTCCTTTG ACCAATAAAT
9541 CGCTTGGTCG AGTGGCGTAA CGGTTTACCG TCTACACTTA TCACTCAAAC CAAACCAAAC
9601 CATCGAAGAA GTGACCTATC GGTTCGAGGG AACGGTGATG TTCTTACGAC CAAGTTAACC
9661 CAAAGAGCGT TCCACATCGT TGAACCGTCT CCTCCAGTTG GATCTGTTTA ACTTCCGCAG
9721 CGACTGAAGA AGGTATCACT TTTTTTTTGG TTCCAAAAAA AAAAAAAAAA ATTAC
```

Figure 8 Nucleotide sequence of the *his-3 cog^E lpl* region of linkage group I in the StLawrence wild type strain of *Neurospora crassa*. This differs from that in the Lindegren strain in many positions, summarised in figure 5. The coordinates of relevant features are given in the text. This sequence contains the weak recombinator *cog^E* and also the remnant of a transposable element *Guest* within the replaceable sequence 3' of *his-3*. StLawrence strains carry *rec-2$^+$* which prevents the initiation of recombination at *cog*.

```
   1 ACCGGGAATC GTAGCGGGCG CTAAGGCCAA GCCGCGGCAC GGGTCACTGA CCCAATGCAG
  61 CGCATTCCGT CAGCAACTGA AGTGGATGTA CAAGTACATA GTAGTAGATC GCAACTGGAG
 121 ATCACTCGCA CCGTGCCGCA GAACAAGGGC GACGAGCCTC AGGGCAGTTT AGCCTGCCGT
 181 AACAGCACAG ACCATAGCTT ATTTTCACCT GGGCGGGCGG GCGACGGCGG CACTGACATC
 241 GGCAAGGCGG CATCAAGCAA CCCCTCTGTT GCTTGCCAGC TGCCGGCCAA CGTCAGCGGT
 301 ACAAGGAGAA ATCTGGAAGG AAAGACTTCT GGCACCGACA GGATGGCACG CGGGAAAAGT
 361 TCCCAATGCA TGAGATGAGG GGCATTTGCA TTGCCTCCCG TCACCCAGTG CGAACCCCAA
 421 CCCCACCATA GCGTCTGTCG ATACATGGAG CGCGAAGTCG AGAAACCTGT AATTCCTGGT
 481 AACTTTCAGG TACACAGTAC GTACTGATCC TGGTATCAAA CCTTGCCTGC CGAGTTTTCG
 541 ACGGAAAGAG GTGTGAATTG TGAAAGAGTC ATACCAAATC ACCCGATTTT CATAAAGCCC
 601 GAGTCTTTTC TGTACATAAG CGACACTCGA AGCGGGCCTC ATCTTCATAG CCTGATAGCT
 661 TGTAATACTC CATCCTCGTA TCTCACTTGA CCTTGAGTTC AACCCCACGT CAAACTTCAC
 721 CCGACACATC GACGGATTGG GGAACAGCAC AATACCTGAA AAGCGAGAAA ACCAAACAGA
 781 GGAAAACACC ATGGAGACAA CACTTCCCCT CCCCTTCCTC GTCGGTGTCA GTGTTCCTCC
 841 CGGACTGAAT GACATCAAGG AGGGCCTCAG CCGGGAGGAA GTCTCGTGTC TTGGCTGCGT
 901 CTTCTTCGAG GTCAAGCCCC AGACCCTTGA GAAAATCCTG CGATTCCTCA AGCGTCACAA
 961 TGTCGAATTT GAGCCCTACT TCGATGTAAC AGCCCTCGAG TCTATCGATG ATATTATCAC
1021 TCTTCTGGAC GCCGGCGCCC GCAAGGTGTT TGTCAAGACC GAGCAGTTGG CCGACCTCTC
1081 CGCATATGGC TCCCGCGTTG CCCCCATTGT CACTGGAAGC AGCGCTGCTT TGCTTTCCTC
1141 CGCCACCGAG AGCGGCCTTT TGCTCTCCGG CTTCGATCAG ACTGCCTCCG AGGCTGCACA
1201 GTTTCTGGAG GAGGCCAGAG ACAAGAAAAT TACCCCCTTC TTCATCAAGC CCGTTCCTGG
1261 GGCCGATCTC GAACAGTTCA TCCAGGTCGC CGCCAAGGCT AACGCCATCC CCATCCTGCC
1321 ATCCACTGGC TTGACAACAA AGAAGGACGA GGCCGGCAAG CTTGCCATCT CCACCATCCT
1381 CTCGAGCGTC TGGAAGTCTG ACCGTCCCGA TGGTCTTCTC CCCACCGTTG TCGTTGATGA
1441 GCACGACACT GCTCTGGGTC TGGTCTACAG CAGTGCCGAG AGTGTGAACG AGGCCCTCAG
1501 GACACAGACT GGTGTCTATC AGAGCCGGAA GCGCGGTCTC TGGTACAAGG GTGCTACTTC
1561 CGGAGACACT CAGGAGCTCG TCCGCATCTC GCTTGACTGC GATAACGATG CTCTCAAGTT
1621 TGTCGTGAAG CAGAAGGGTC GTTTCTGCCA CCTCGATCAG TCCGGCTGCT TTGGTCAGCT
1681 CAAAGGCCTT CCCAAGCTCG AGCAGACTTT GATTTCGAGG AAACAGTCTG CCCCCGAGGG
1741 CTCCTACACT GCCCGTCTCT TCTCCGATGA GAAGCTAGTC CGGGCCAAGA TCATGGAGGA
1801 GGCTGAGGAG CTCTGCACCG CTCAGACCCC CCAGGAAATC GCCTTTGAGG CTGCCGATCT
1861 CTTCTACTTT GCTCTTACCA GGGCCGTTGC TGCCGGCGTT ACTCTTGCCG ATATCGAAAG
1921 GAGCCTTGAC GCCAAGAGCT GGAAGGTCAA GCGCAGGACT GGAGATGCTA AGGGTAAGTG
1981 GGCTGAGAAG GAGGGCATCA AGCCTGCGGC GTCCGCTCTC GCTGCCACTT CGGCCCCTGT
2041 CACCAAGGAG GCCGCCCAGG AGACCACCCC TGAGAAGATC ACCATGAGAC GTTTCGACGC
2101 CTCCAAGGTC TCTACCGAGG AGCTCGATGC TGCTCTCAAG CGTCCTGCGC AAAAGTCGTC
2161 CGATGCCATC TACAAGATCA TTGTCCCCAT CATCGAGGAC GTCCGCAAGA ACGGCGACAA
2221 GGCTGTTCTG TCGTACACTC ACAAGTTCGA GAAGGCTACC TCTCTTACTA GCCCCGTCCT
2281 GAAGGCGCCC TTCCCCAAGG AGCTTATGCA GCTCCCTGAG GAGACCATTG CTGCCATCGA
2341 CGTGTCCTTC GAGAACATCC GCAAGTTCCA CGCCGCCCAG AAGGAGGAGA AGCCCCTCCA
2401 GGTCGAGACC ATGCCCGGTG TTGTCTGCAG CCGTTTCTCT CGTCCCATCG AGGCCGTCGG
2461 CTGCTACATC CCCGGCGGTA CCGCCGTTCT CCCCAGCACT GCCCTTATGC TGGGTGTTCC
2521 CGCCATGGTC GCCGGCTGCA ACAAGATTGT GTTCGCCTCT CCTCCCCGCG CCGACGGAAC
2581 CATCACTCCC GAGATTGTCC ACGTCGCTCA CAAGGTTGGG GCCGAGTCCA TCGTGCTTGC
2641 CGGCGGTGCC CAGGCCGTAG CTGCCATGGC CTACGGCACC GAGAGCATCA CCAAGGTCGA
2701 CAAGATTCTC GGCCCCGGTA ACCAGTTCGT CACTGCTGCC AAGATGTTCG TCAGCAACGA
2761 CACCAACGCT GCCGTTGGTA TTGACATGCC CGCTGGCCCG TCCGAGGTGC TGGTCATCGC
2821 TGACAAGGAC GCCAACCCCG CGTTCGTTGC CTCGGATCTC CTGTCCCAGG CTGAGCACGG
2881 CGTTGACAGT CAGGTCATCC TGATCGCTAT TGACCTCGAC GAGGAGCATC TTCAGGCTAT
2941 TGAGGACGAG GTTCACCGTC AGGCTACGGA GCTTCCTCGC GTCCAGATTG TCCGTGGCTC
3001 CATCGCCCAC TCGATCACCG TGCAGGTCAA GACCGTCGAG GAGGCCATGG AGCTCAGCAA
3061 CAAGTACGCT CCTGAGCACT TGATCCTCCA GATCAAGGAG GCCGAGAAGG CTGTCGATCT
3121 TGTCATGAAC GCCGGTAGTG TCTTCATTGG CGCCTGGACT CCTGAGTCCG TTGGCGATTA
3181 CTCTGCTGGT GTTAACCACT CGCTGCGTAA GTTACATATC ATAAATAGCC CCGCTTCACA
3241 GATTCTTCTG CTAACGTCAA GACACATAGC TACCTATGGC TTTGGCAAGC AGTACTCTGG
```

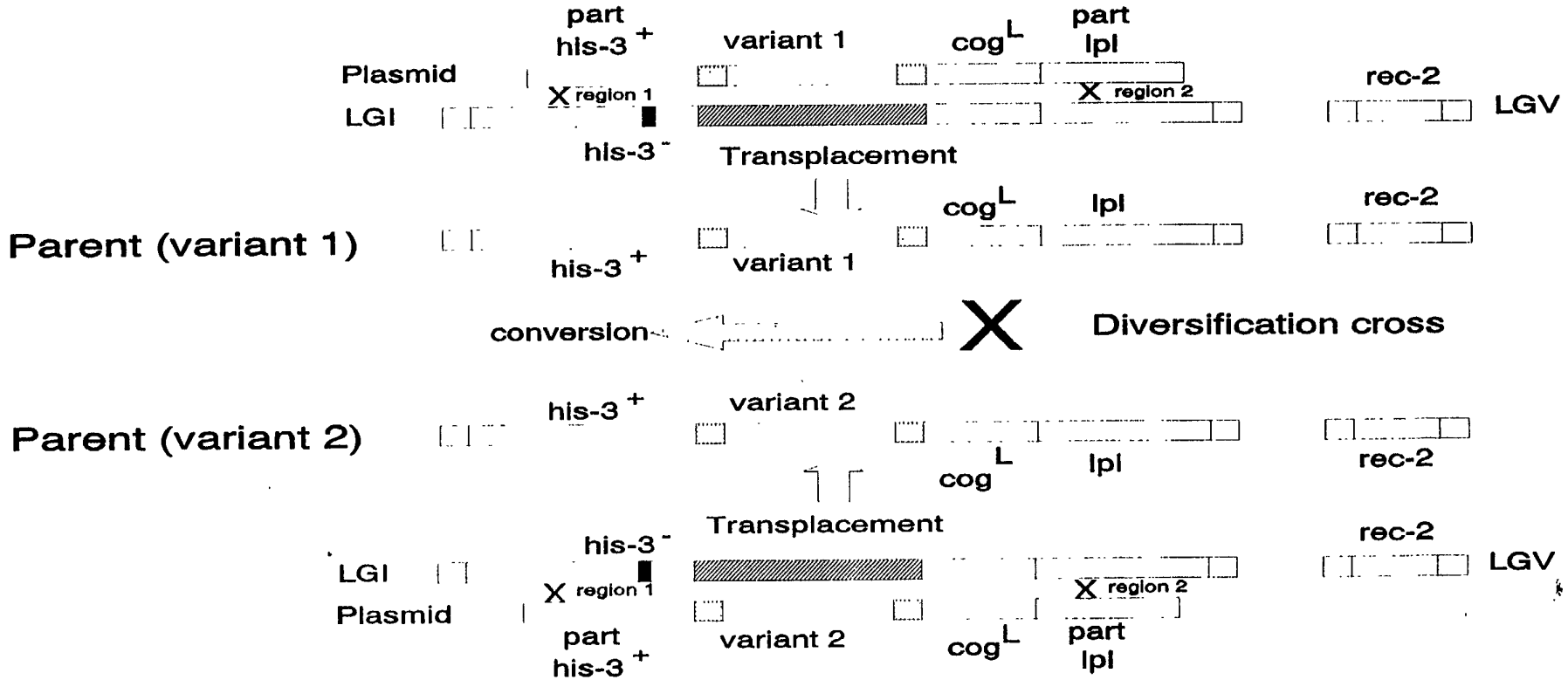
Figure 8 continued

```
3301 CGTCAATTTC GCCTCGTTCG TCAAGCACAT TACCAGCTCC AACTTGACTG CCGAGGGTCT
3361 CAAAAACGTC GGCCAGGCTG TCATGCAGTT GGCTAAGGTT GAGGAGCTCG AGGCTCACAG
3421 AAGGGCGGTC AGCATCCGTC TTGAGCACAT GAGCAAGAGC AACTAAACGG AAATTCTTTT
3481 CGAAGTAGCA AAAAAAAAAA AAAAAAACAA GAACAAAAGG ATGTAGTGGG TTGATGTATA
3541 TCTGGGTCAT TTTGGGCACA TAGAGTAATG ATAACGAGTT TTGGACATTG TACTGTTCTG
3601 TACAGGCTGA AGATCAGTAC ATGAATCTGT TGGTAAGTGT GGAGACCCAA ACGTCCCTTG
3661 AGTTTTTCTC CCTATTCCAG AGGTGCTCGT CCCTGGGTGT TTATTTTCAT TATTACATCA
3721 ACCTTTTTTT TTTTTTTTTT TTTTTCAGAT CATGCGTACA TGAACGGGGG AAGCACAGAC
3781 GATCGAAACG TGGATGTCAC AATGTCGCTG CAGTGATGCT GCATTGCATG AAGCGCCCAT
3841 CTCAATATAC TTGCAGTCTT GCACGTTGCA TGTGAACTTC CCAAACAACC GAATAAAAGA
3901 CGGCGAAAAA TGAAGATAAA AAAAAACCAT AAAAAAAATC AGAGGGAGTG TGGGAAATGG
3961 TGTCTTTTAG CATTCAGACC CCATAGCCGT GCACGCCCGG GTACAGACAG GTTCATCGAT
4021 GTTGACATTG ACTGGGACAC CAGGTCTATC TATTTTATCT CCTGTCCTCT ACCATACATC
4081 GGGACATCGG ACATCTTGCT GTACCCCCCA CACCCACAAA GCCTTATAAA AGCGCCACAC
4141 CCGAGGAGGT TCGGTCGGCC CCACGAACTC TGTGCCTCCC TGCCTGTTTA CAGGGACCGA
4201 ACGCTGGAGA ATCTTACTAG TTTCCTGACA TCCGGCCTAC CCGAGCAGGA AAAGGGACAG
4261 CTCATAGGCG AGGAGGGATT TGAAGATGGG AACATTTTGG GTGATTCGAG AGGAGGAACT
4321 AGGTACTGCA TCATGATAGT TCGGGGCAGC ATCTTGGCTG GGACATTGTT AATACCTCGA
4381 TATGATGAAG TAGGAGGGAG TTTTTGCGTG TCTTGCCGAA GTCCAGAGAT CTGTTTTATT
4441 TTATTTTTTA TGGATGTAGT GTATCAACAC CCAAGATTCG GAGAATAGTA CTAGGATTCG
4501 CATTTACAAG TGGAAGTCTT GAGAATCGTT GTATATCCTT GTCTTCCTCG GAATGTTAAC
4561 AATCCTACAG CGAGCGAGCG AGCGGTCGGA TGCGCTGATC TGATAGGCGC AATATACGGC
4621 CGCTTTCTCC GGTCGTGTAG TGTAAGCTCT GTGGGCATAG TACACTAAAA AAACCCTTGC
4681 ATTTCATGAT CTGCCTGCTA TTCATTCCGA GCTATTTCAG TGGTCACATT TCGAGGAAGA
4741 AAGAAAGCAA CTAAGATTCA CAGCCATCCA TCCATCCATA TGGAAGAATA ATCCATTCCC
4801 ATGTTCCCTC CCCCCCACTA TGTATGTGAC CACACGCTGC TGTCAGAATG CCAACGGTCT
4861 CAGGTACCCT CGTCCGACTG TTTGGCATGG AGTTACATAC ACTACTAGTG TAGCCCCGGG
4921 CCAAGCTACC CCGTCAAATC TATACATATC TATAACGGGT TTCAGGGGTT TCGTTCGCTG
4981 TCAATCAAGT TTGAAACATC ACTGGGGCCG TTGGACGGTG TATTAGACCA TTGGCTCCCT
5041 CAGCTGTTTG GCGGCTGGGC GGCTGGGTCA AACGGCAATA ACGGGACTCG AGAGGGACGA
5101 GGAGAGTCGG TTGGCTGGCT GCAATACAAG CGTTCCCACC TAACCAACGA GTCCCGTTTT
5161 CCATTTGTGT GCCTAACCAT CATCTAGGGA TGTCAGGGTT TGGCCGGATC AGGGTATGTT
5221 TGGTTGACTG TTGTCATGTC TGATTGGGTA CATATTATGG TAGGTGTCTC GAGAACAGTA
5281 GAGTACTCGG GCCTAGCGTT TGGATGATTA CGCGAGATAT GAGTTGTGGG CCGCCATGCA
5341 GTTGCTTGTC CATAAGCAGA AGTTGCTTTG GGATATATTT CTCGTCTTTC AAAGGTCACG
5401 AGGTCCTGGG ACGAACGGCA TCGCCATCCA AAGGGTTGAA CATGAGAAAC CTGAATGGCC
5461 TTTGCGTTGA AATACAAAAA GTCAAGAACA AAATCGCTTG AGGATAGGGA CGTGGAAGCA
5521 AGCAAATATG GTAAGAGAGG TATACATCAA CCCTGGTTCA ATTGTTAGCG TGGTTCTTCC
5581 TCCACGTCCT CGTTCATGAC GGTTAACAGT ACCAGGCTAA CAATTAAACC AGGGTTGATG
5641 TGTACTGATA TGTAGGTGCT CAGCAAACTG CCAATTTCTT TGGCCCCAAG CAGCAGTTTG
5701 CTGTCAGTGC TGCTCGTGTC AGCCTTGGTA GTGGAACCTA AACTGCTAAC ACAGCGCAAG
5761 TGCGCATGTA AAGATATTGT GGGAGGATCT GTATGGATGG ATGAGATTAC TGCTTGGTGT
5821 TGGTTGCGAG GCACTGCGGC TGTTAGGCTT TGCTGTGCCC CGTTCGACGA AGAAATACGC
5881 GGAACTATAA ATTGGATACC TAGACTTACT GCCTATGGGA GGTATCTACC GACGTAGCCG
5941 ACGGATTCTA GCAACATCCC GACTTTGCTT GTAGTGTACT ATGATAGCAG CACAGTGTTG
6001 CTCCTTGTGA GAATGGGCTC TTTTTTTTTT TCCCCCTTCC CTAGGGCGTT GACTGGACTT
6061 GCTCTATTGT TCCCAAGGTA GGTGCCCGTC ATCGATTTTC CCAAGTCTCC CGCCAGATTG
6121 TCGTCATAGT GTCATGATGA CCTCGGTCGC TGGGGCTGCG TGGTTACGGG GAGCTGGGAC
6181 CGCTAGGCCT CAGTGGTTGT GCCATTCAGC GTGGGTGTGT GGAGTAGCGG TAGAGGCGCT
6241 TGGAAGTTGT GCTAGCGGAA ACCCTGGAAT ATCTTCTACC CTCGATTCCT TCTCGGGCTG
6301 CCCATGTGCT GAGGTGATGC CGGGGATCTG GCGCCAATCA TCCATTGAGG TTCCCGCAGC
6361 TTCCCGGTGC CGCGCGCGGG CGCAGTTGCT CACAGGACAC ACCTAGACGC AGGGGCACAG
6421 GGGCACCGTT TGGTGTGCAA CTGGGTACCT AGCTGTAGCA AGCACTCCAC CGTCTGTGCA
6481 ATCCCCCAAT CCACGGCAGG AACTTCGCAC CGCCGCGGCA CCGAGTGAGC GAATCCATCC
6541 GCATTGGATC CCAATTCTTG CCCTTGCCAT CCTTCTTTCT TCCCACTTGG CGCAACCAAC
6601 ACTTCCCTTG GTCTGGGTAC TCGTGTTGAT CTTCACTCTC TTTTTTTCTT GGGCGACCGA
6661 CTTTTTATAT CCGTCCTTGC TTCCCCCTGG CCGTTGTCGT TCTTTCTACA ACTACCTTCC
6721 GTTCATTATC CCCTTTCTTG GTTCGGTCGA GGACCCAAAA ACAGAACAAT TCCGGCTCTT
6781 CCAGGTGGCT TGGGTGCGAC TGTTTAGCTC TTGACCACTA GCCGCTTACC TTCTCTTGAT
6841 GTTTTTATTT GGATATCATT AAACTACTCT TTCTTGAAAC GGCAGACGAA CGGAACAGTT
6901 CCTACGGTAT ATTAGCGATA TACGTTGTAC TGATATTCTG AGCAAGAAGA GGCAAATTAT
6961 CAATTATGCA TCTCCCTTCG TCGCTGCTCA TCGCAGCTCC CTTGCTCGCC AATGTATCGG
7021 CCGAACCCAT TAGGATACCC CAACGCGATG TTCTCCGTGG TATCAACATC ACAGCAACTT
7081 GCCGTTCGAG CACTACCGGA TTCGCCCAGC GGTGGATATG CCCCTGCCGT TGTAGACTGT
7141 CCCAAGACCA AGCCGACGCT CCGGAAGGCC GTGGATTTGT CGAACGAGGA GAAGAACTGG
7201 TTGTCGATCC GGAGGAAGAA CACCATCCAG CCCATGAGGG ACCTCCTGAA GAGGGCCAAC
```

Figure 8 continued

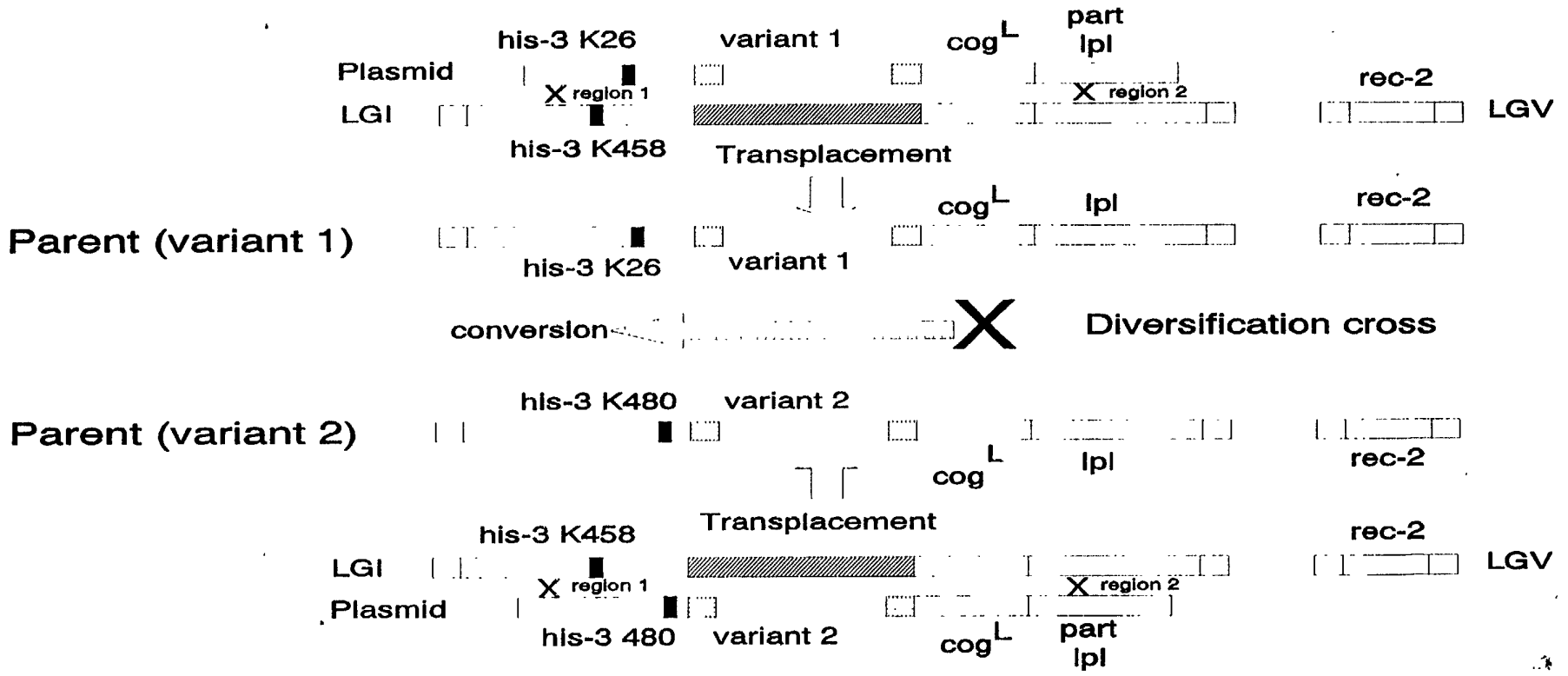
```
7261 ATCACTGGGT TCGATTCCGA GACATTTATG AATGAGGCCG CCAACAACAT CTCGCAACTG
7321 CCCAATGTCG CCATTGCCAT TTCAGGAGGC GGCTATCGTG CCCTCATGAA CGGCGCCGGC
7381 TTCGTTGCTG CTGCGGATAA CCGAATTCAA AATACCACGG GCGCAGGTGG TATTGGAGGC
7441 TTGTTGCAGT CCAGCACATA TTTGTATGTA AAGTGGTTCT TCTTATCTCG TTTTCGAGTG
7501 TCAACTGCGC CAGTTCAGAG TTGGGCGGCT GTGGACGACC TTGCTGGTGA ACATGTCTTG
7561 GACTCCATGC CCCTTCTTCG TTTCCTCAAA TCAAGAAGTC GAGGACCGTG ACCGTAAATC
7621 GCTAACGCAA CTCTAGGGCC GGACTTTCTG GTGGTGGCTG GCTTGTCGGC AGTTTGTTCT
7681 CCAACAACTT CAGCAGCATT GAGACCCTGC TGAGCGAGAA CAAAGTCTGG GACTTTGAGA
7741 ACTCCATCTT TAAAGGGCCC AAGGAGGCTG GCCTTAGTAC TGTCAACCGC ATTCAGTACT
7801 GGTCCGAAGT GGCAAAGGAA GTTGCCAAGA AGAAGGATGC TGGCTTCGAG ACAAGTATAA
7861 CAGACTACTG GGGCCGAGCA TTGAGTTACC AACTGATCGG AGCCGATATG GGCGGCCCGG
7921 CTTACACCTT CTCCAGCATT GCCCAGACCG ACAACTTCCA GAAGGCCGAA ACGCCGTTCC
7981 CTATTCTGGT AGCTGACGGC CGCGCGCCTG GAGACACCAT CATCTCCCTC AATGCTACCA
8041 ACTACGAGTT CAACCCGTTC GAGACGGGTA GCTGGGACCC GACCGTCTAT GGCTTTGCGC
8101 CGACCAAGTA CCTCGGCGCC AACTTCAGCA ACGGCGTGAT CCCATCGGGA GGCAAGTGCG
8161 TTGAGGGTCT CGACCAAGCC GGCTTCGTCA TGGGCACCAG CAGCACGCTC TTCAACCAGT
8221 TCCTTTTGGC CAACATCTCC AGCTACGACG GTGTTGCCCG ACGTGCTCAT CGAAGCCGTG
8281 ACTTCTGTCC TCAAGGAAAT CGGCGCCAAG AGGACGACGT CTCCCAAATC ATCCCTAATC
8341 CGTTCCTGGA CTGGAACAAC CGGACCAACC CCAACGCCGA CACGCTCGAG CTCGACCTGG
8401 TCGACGGCGG CGAAGATCTG CAGAATATTC CGCTCAACCC GCTCACCCAA CCCGTGCGCG
8461 CCGTGGACGT CATCTTCGCT GTCGACTCGT CCGCCGACGT GACAAACTGG CCCAATGGCA
8521 CCGCCCTGCG AGCCACCTAC GAGCGCACTT TCGGCTCTAT TTCCAACGGG ACACTCTTCC
8581 CCTCGATCCC CGACGACTGG ACGTTTATAA ACCTAGGCCT CAACAACCGC CCCTCTTTCT
8641 TCGGCTGCGA TGTTAAGAAC TTTACCTTGA ACGCCAACCA AAAGGTTCCC CCCTTAATCG
8701 TCTATGTCCC CAACGCGCCC TATACCGCGC TGAGCAACGT GTCCACCTTC GATCCGTCAT
8761 ACACCATGTC TCAGCGCAAC GACATCATCG GCAACGGATG GAACTCAGCC ACGCAGGGAA
8821 ACGGCACGCT GGATTCGGAG TGGCCCACTT GCGTCGCCTG CGCGGTTATC AGCAGGAGCT
8881 TAGATCGGTT GGGCAGGCAG ACGCCAGCCG CGTGCAAGAC TTGCTTTGAG AGGTATTGCT
8941 GGAATGGCAC AGTGAACTCA AAAGATACAG GGGTTTACAT GCCTGAGTTC AAGATTGCGG
9001 ATGCGCATGC CCTGGACTCG GGTGCTGTTG CTATCGGAAA GATGGTGAAT GTCTGGTCGT
9061 CGGTTGTGGT GGGAGTTGTG GCGGCTACTT TGTTGTTGTA GGGGTAGGGG AGACGTGATG
9121 ATATTCCAGT CTGATGAAGT TGAGACTGGA CTGGAGATCG CCAAGGATGC GGAGGGAAAG
9181 GAATGCGTGG TGTTAATGTC ATGATGGATG AAGGGTCATG GATCATGGAA CGACGGGGCG
9241 GGGATATTGG ATGATGGATA TACCACACTG CATGCATGCT CTATTGATAA TATGCTTTGG
9301 CATTTACGTT TAACAATCAA TTGCTCCATC CTGATGTTCT ATCTTTCGAC ACTGGATTGA
9361 TACTACTCCT GTTGCTTCCC TCTTGAAGTT GGAAGGACTT GAGGTTGGAA GGACTTGAGG
9421 TTGTTTGTTC TGAGGGAGGT TATCGAAGTA TCATCTGTGC TGATGCCGAT CGATAGACTG
9481 CCCTCTTCTT CGAGGCAACG AACGGTCGGA TGAGCCTCTA ATCATGATGC TCAGTGCCAC
9541 AAAAAGGCTC CAGCACAGCT GCCCACACCT TTTTTGCCTC GTCGCTCCTT CCTTTTTTTC
9601 CCCCCCTTTC TTCCTTTCCA TCTCATCCCG TACCAGAGTG CCCACCGGGT ATATATATTA
9661 CCTCCTTGGC CGTTCTCCTT TGACCAATAA ATCGCTTGGT CGAGTGGCGT AACCGTTTAC
9721 CGTCTACACT TATCACTCAA ACCAAACCAA ACCATCGAAG AAGTTACCTA TCGGTTCGAG
9781 GGAACGGTGA TGTTCTTACG TTCAAGTTAA CCCAAAGAGC GTTCCACATC GTTGAACCGT
9841 CTCCTCCAGT TCTTGGATCT GTTTAACTTC CGCAGCGACT GAAGAAGTAA TCACTTTTTT
9901 TTTTTTTGGT TCCAAAAAAA AAAAAAAAAA TTAC
```

Figure 9 Construction of the components of the sequence diversification cross: Parent (variant 1) and Parent (variant 2). For convenience, plasmid sequences are shown as linear. The cross hatched region in the chromosome is dispensible. Stippled sequences in the plasmid indicate the multiple cloning site for inserting foreign DNA. Crossovers in region 1 and region 2 insert the foreign sequence to be diversified into chromosome I of *Neurospora crassa* adjacent to the recombination hotspot *cog*. Parent (variant 2) containing a version of the foreign sequence with multiple differences from that in parent (variant 1) is similarly constructed. Parent (variant 1) and parent (variant 2) are crossed and conversion events (stippled arrow) initiated (X) at cog^L recombine the sequence differences in variant 1 and variant 2 to form new combinations. Sequences are identical except for those that distinguish variant 1 and variant 2. *rec-2* on linkage group V permits cog^L to be active. For simplicity, genes not directly related to the diversification are omitted. See text for further details.



006231 29615760

Figure 10 Construction of parent (variant 1) and parent (variant 2) enabling selection of progeny that have experienced conversion in the foreign DNA. Complementing pairs of *his-3* alleles are used to obtain parent (variant 1) and a different pair of complementing *his-3* alleles are used to obtain parent (variant 2) as explained in the text. Parent (variant 1) and parent (variant 2) are crossed and his$^+$ recombinants are selected. These must all have experienced conversion events affecting the foreign DNA since the events begin at *cog^L*. The *his-3* alleles in parent (variant 1) and parent (variant 2) are non complementing to ensure that selection yields recombinants and not aneuploid progeny having two copies of all or part of linkage group I.



006237 "396T5260

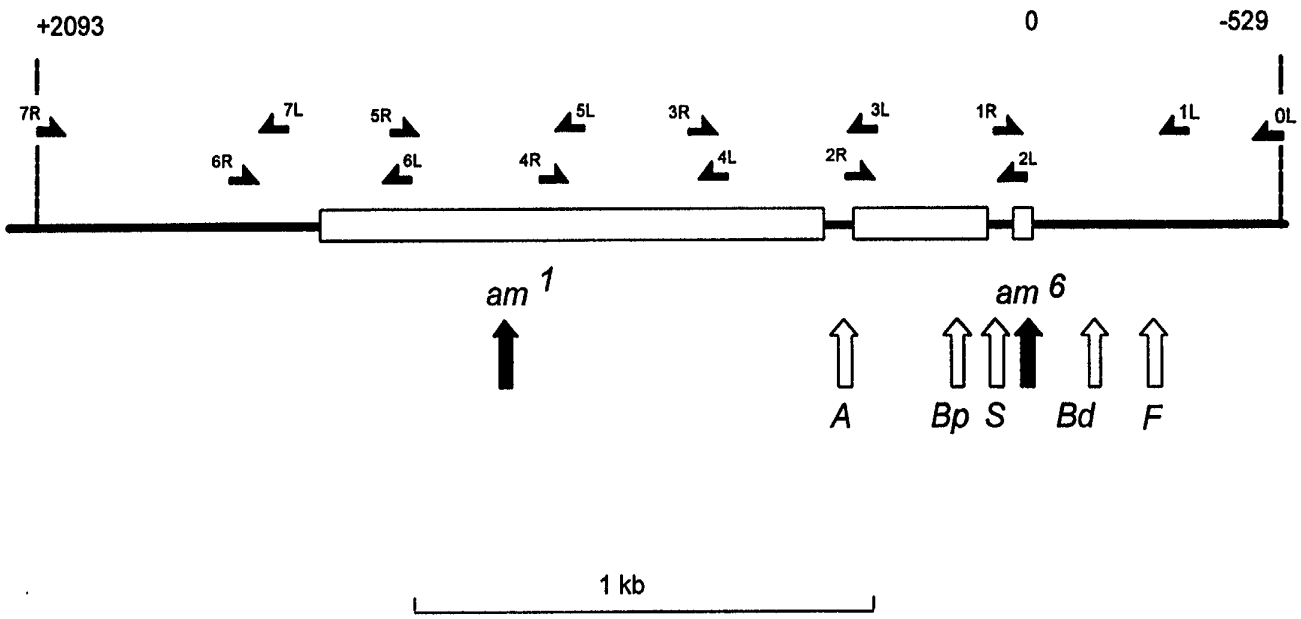


Fig 11

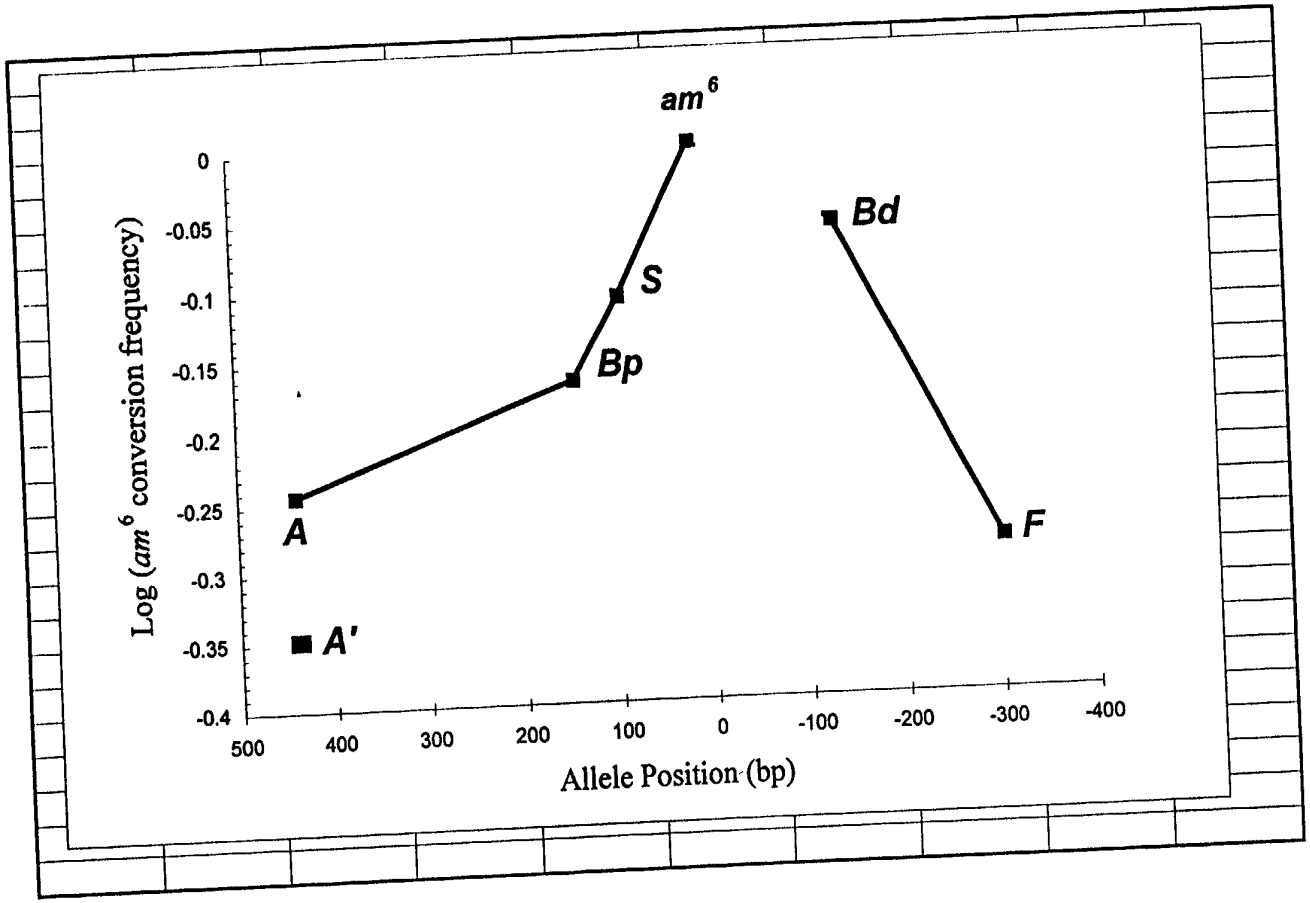

am 6
Bd
S
Bp
A
F
A'
0
-0.05
-0.1
-0.15
-0.2
-0.25
-0.3
-0.35
-0.4
Log (am 6 conversion frequency)
500
400
300
200
100
0
-100
-200
-300
-400
Allele Position (bp)

Fig 12

0962271 3961562

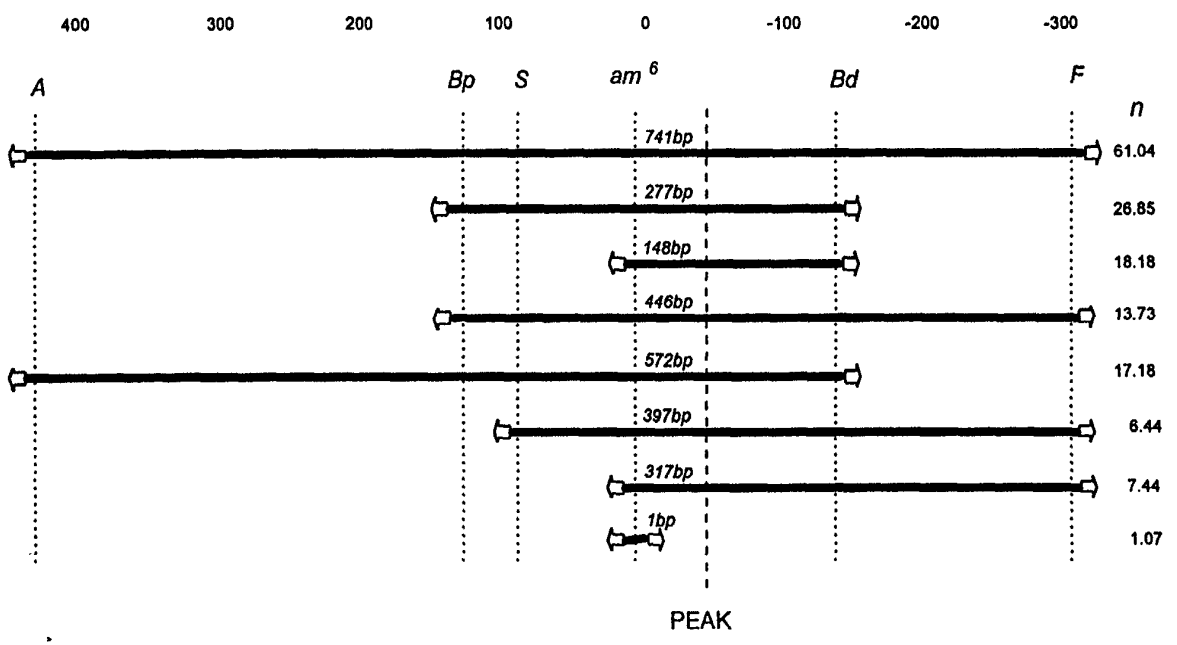
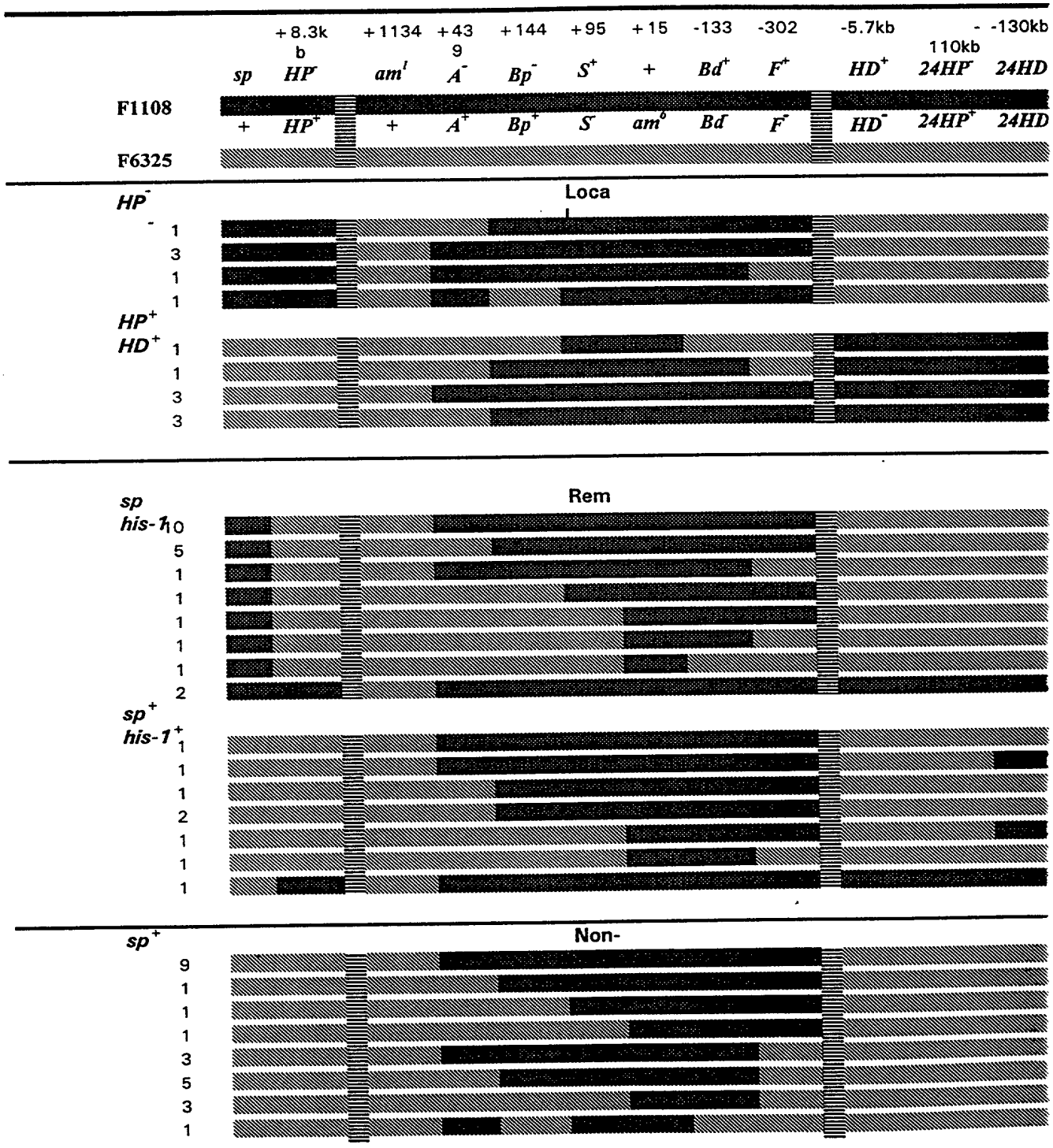


Fig 13

Fig 14 (2 pages)



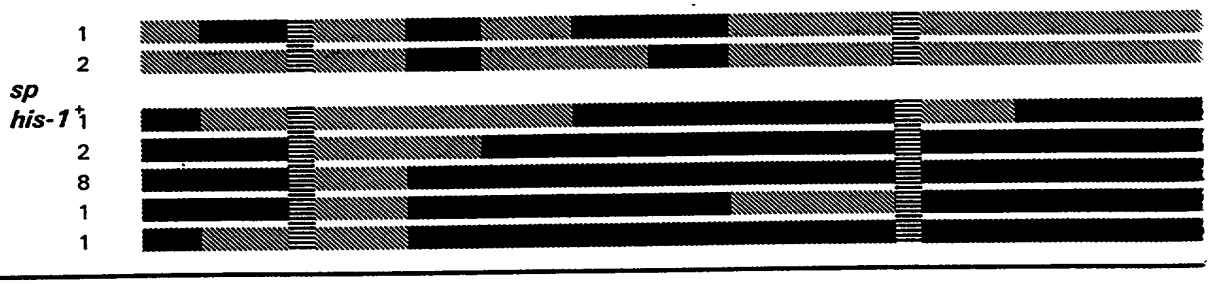


Fig. 14 (continued)

FIG. 1A

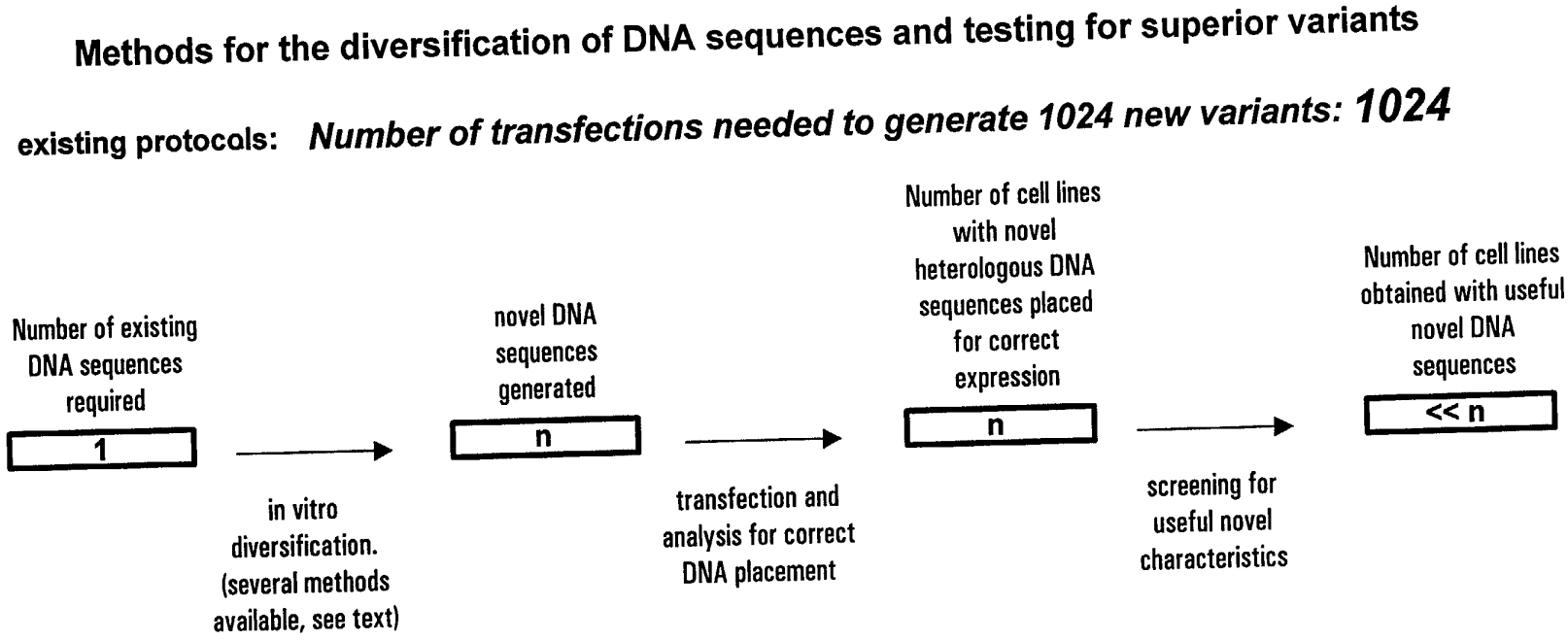

Methods for the diversification of DNA sequences and testing for superior variants
existing protocols: Number of transfections needed to generate 1024 new variants: 1024
Number of existing DNA sequences required
1
in vitro diversification. (several methods available, see text)
novel DNA sequences generated
n
transfection and analysis for correct DNA placement
Number of cell lines with novel heterologous DNA sequences placed for correct expression
n
screening for useful novel characteristics
Number of cell lines obtained with useful novel DNA sequences
<< n

FIG. 1B

A protocol enabled by the present invention: ***Number of transfections needed to generate 1024 new variants: 2***

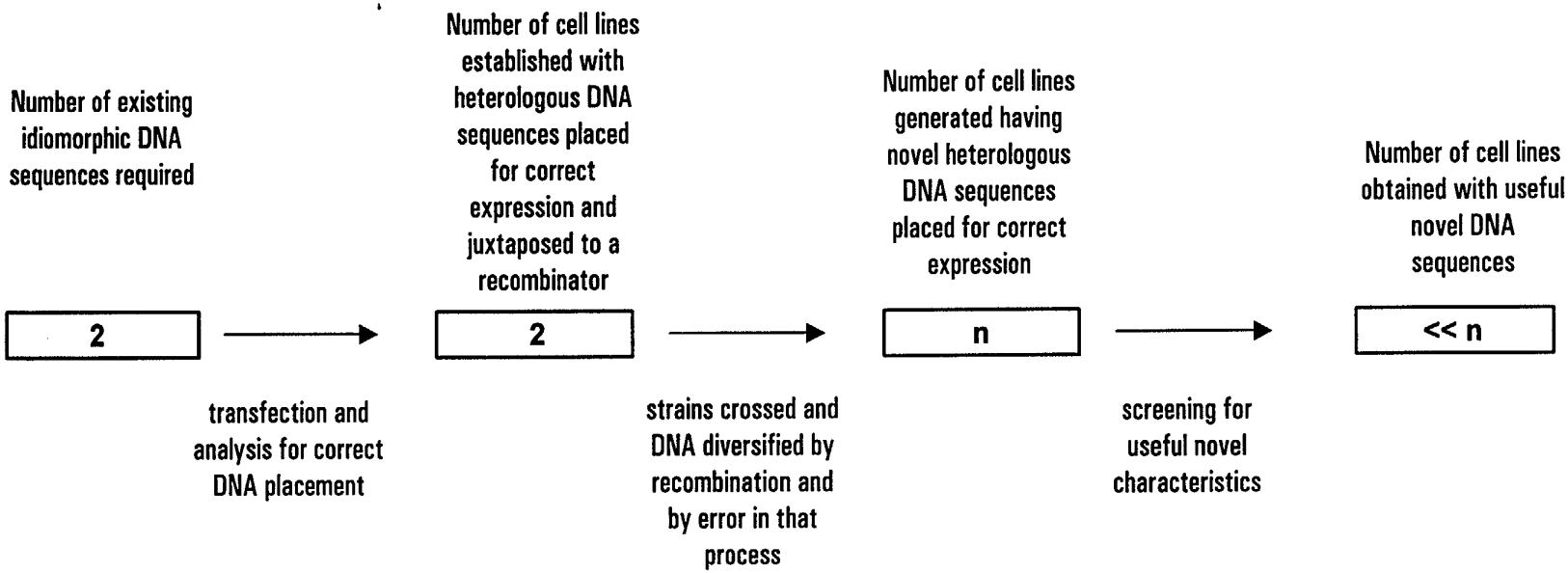


FIG. 2A

Methods for the diversification of DNA sequences coding subunits of heteropolymeric proteins and testing for superior variants.

The example given for immunoglobulins is for illustrative purposes only and is not intended to limit application of the present invention to this specific heteromeric protein. H = heavy chain genes, L = light chain genes

Existing protocol: ***Number of transfections needed to generate 1024 new combinations:*** ***2048***

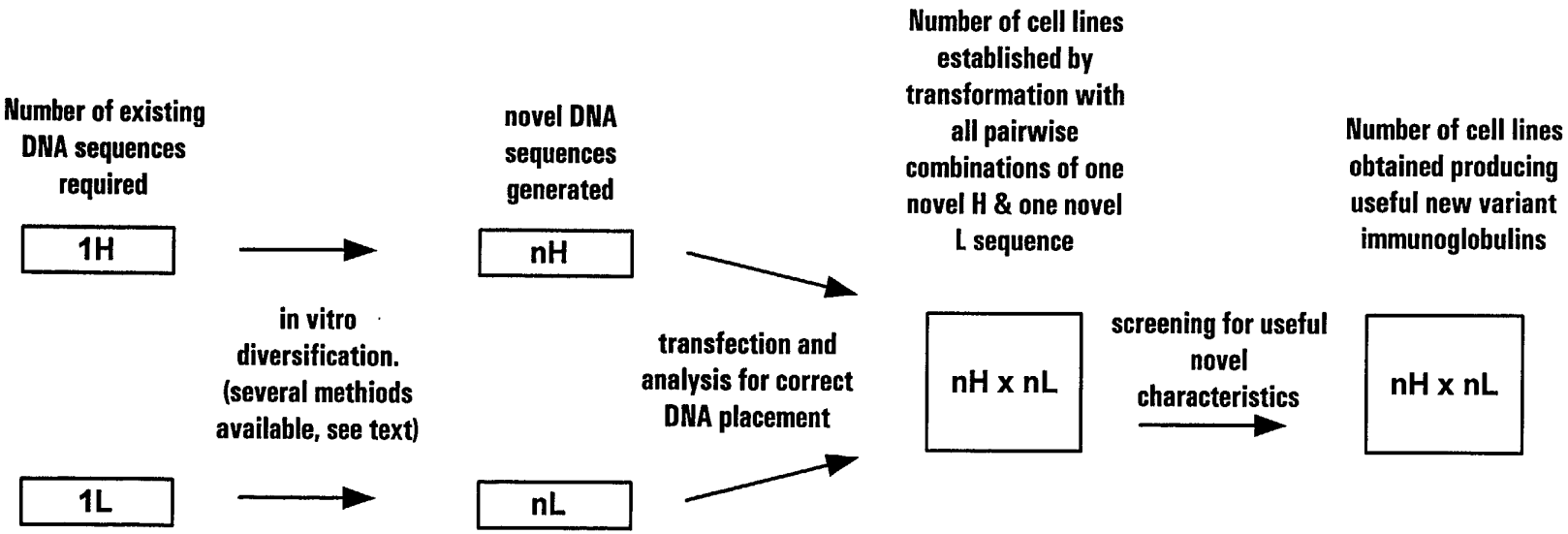


FIG. 2B

Existing protocol using the heterokaryon technology of US Patent Serial No. 5,643,745

Number of transfections needed to generate 1024 new combinations: 64

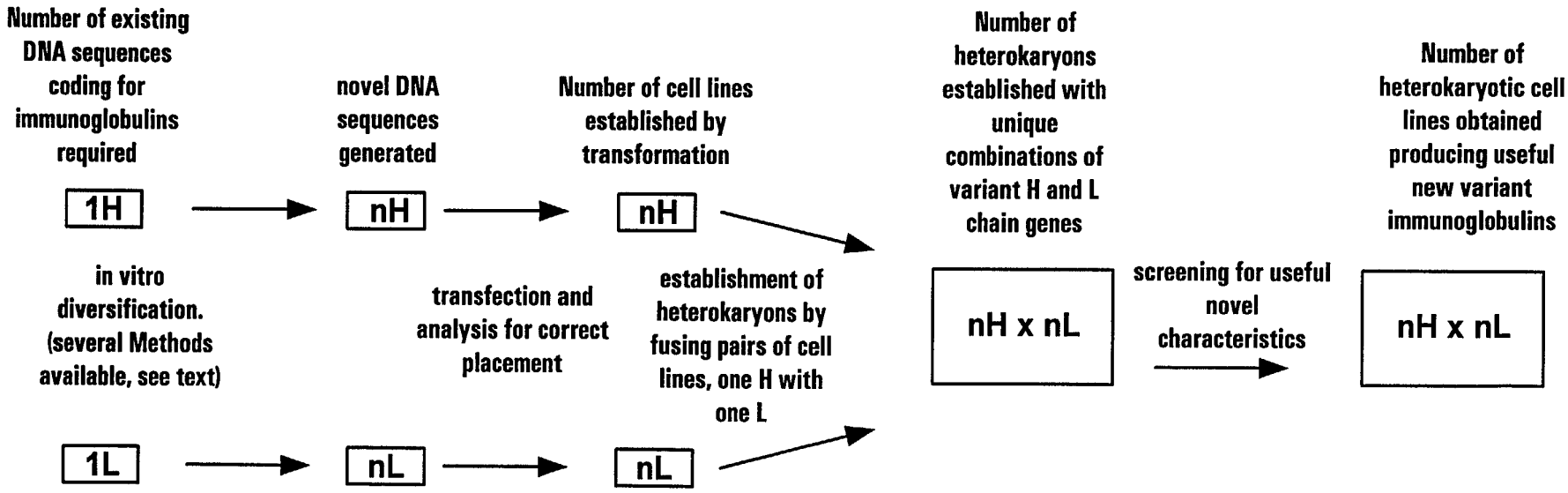


FIG. 2C

A protocol enabled by the present invention

Number of transfections needed to generate 1024 new combinations: 4

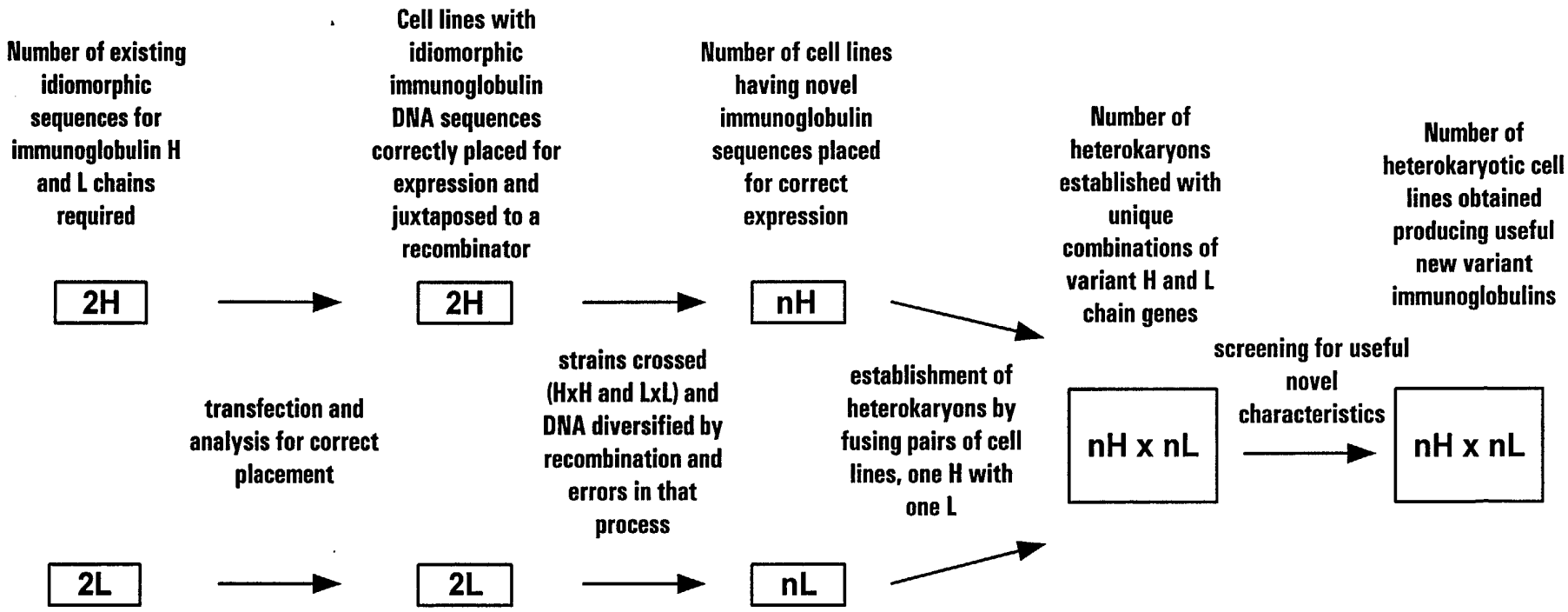


Fig. 3

The modified double strand break repair model for meiotic recombination. After H Sun*et al* Cell **64:** 1155-1161, 1991

(a) A double strand break (DSB) is made in one DNA duplex. (b) A long 3' overhanging single strand tail is generated either side of the break by resection. (c) One 3' end invades a homologous duplex forming a D loop. (d) the D loop is enlarged by repair synthesis and anneals to the second 3' end (e) Repair synthesis occurs at the second 3' end and two intermolecular junctions (Holliday junctions) are formed. Resolution of the junctions by cutting inner and outer strands can give rise to non-crossover (f) and crossover (g) chromosomes. If there are base mismatches in the heteroduplex regions (duplex molecules with thick and thin lines) there will be gene conversion. If mismatch repair does not occur there will be post meiotic segregation of new sequence combinations.

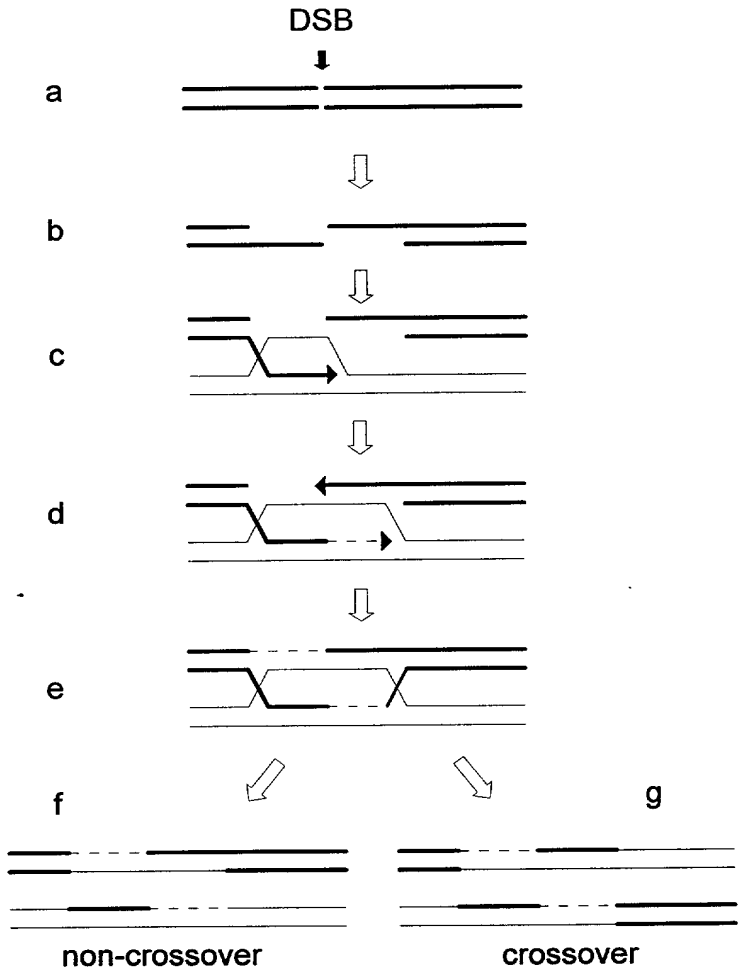
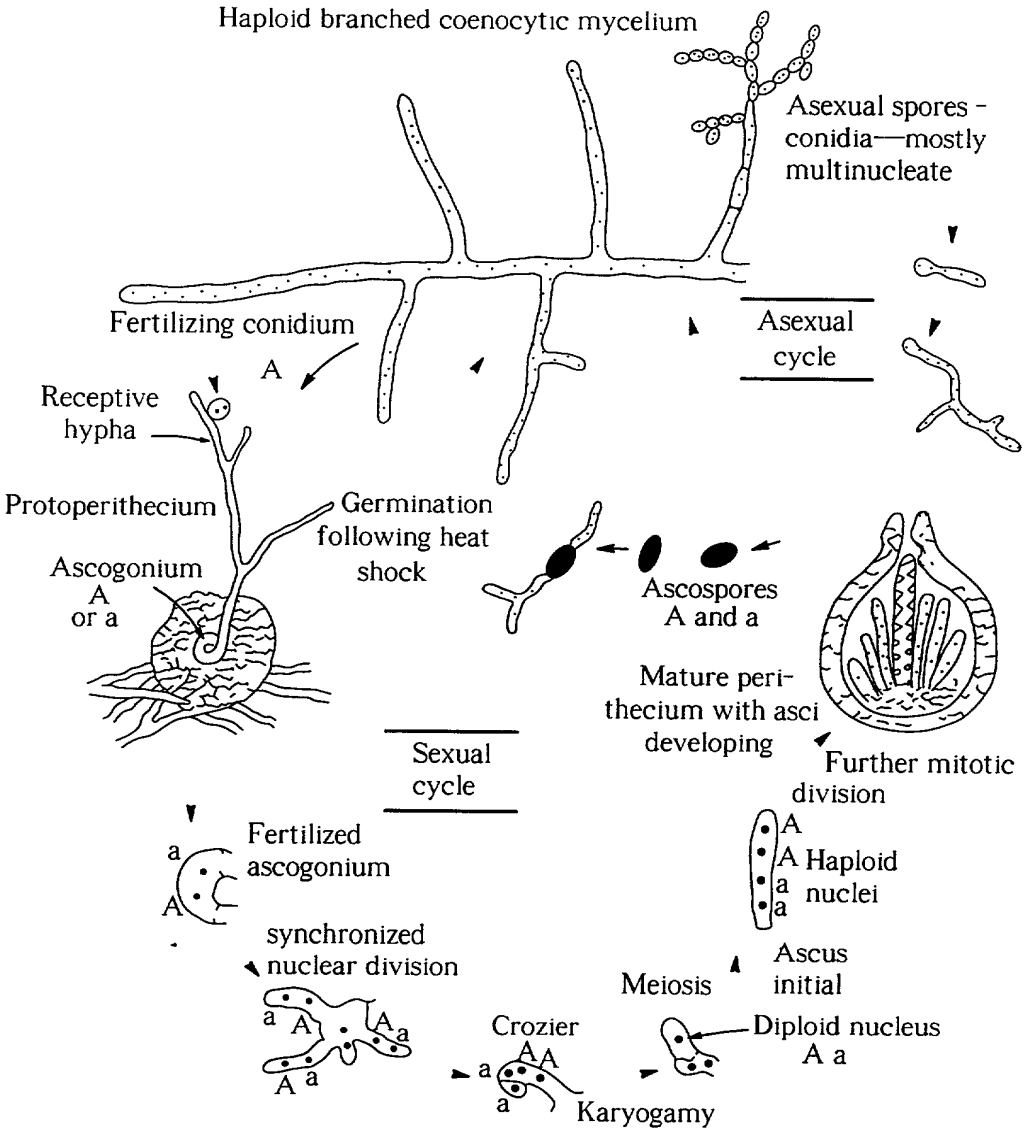


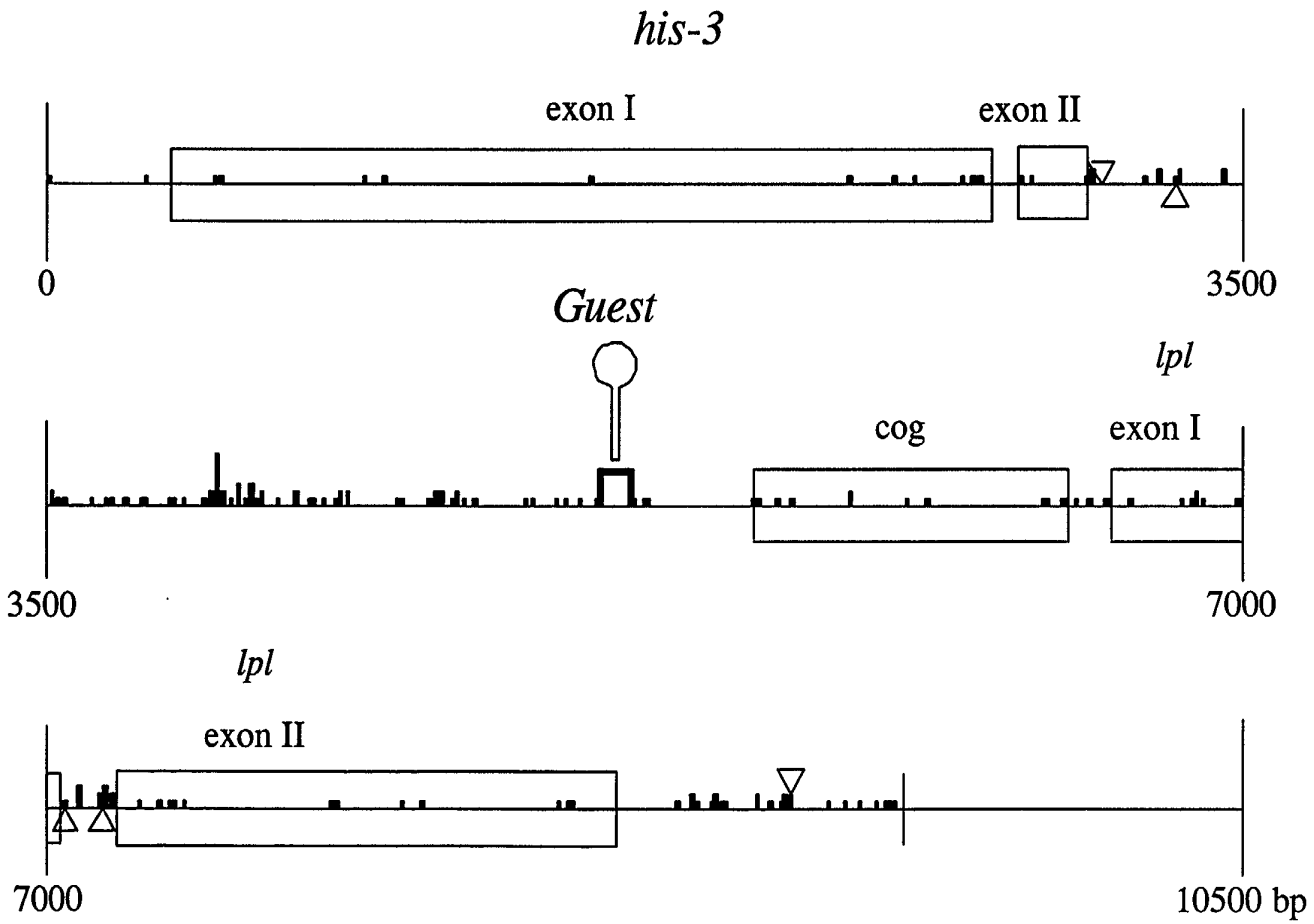
FIG. 4

Life cycle of *Neurospora crassa* after JRS Fincham (Genetics, Wright 1983). Microconidia having one nucleus are not shown but can be generated as described in the text. Perithecia and protoperithecia are shown in section.



006337 3567260

FIG. 5 Map of the *his-3, cog, lpl* region of Linkage Group I of *Neurospora crassa* Vertical bars, triangles and hairpins show the location of sequence differences that distinguish the St Lawrence and Lindegren wild type strains. The corresponding full DNA sequences are given in Fig. 7 and Fig. 8. Vertical slashes indicate one to seven base substitutions per 10 base pairs. Triangles indicate short sequence insertions and the hairpin a 101 base pair inverted repeat transposon fragment present in St Lawrence.



006237 "3967S260

FIG. 6 Discontinuity in the parental origin of DNA sequences in progeny from crosses between pairs of ***his-3*** alleles. In most cases this reflects discontinuity of conversion tracts, in some cases crossovers near the ends of conversion tracts.

Markers are specific DNA sequence differences that distinguish the parents. These were all E (Emerson wild type origin) or all L (Lindegren wild type origin) in the parental strains. Recombinants carry both E and L markers. Marker position is given in base pairs from the first base of the first codon of the ***his-3*** gene. Each line of the table shows the parental origin of the markers inherited by one of the progeny.

marker	P	H	P1	K1201	K504	L3	R1	K26	K874	R4	C4	C5	C6	C6/7	C7	C8	C1	C2	C3	C9	D
location		~6000	−384	115	179	563	1232	1502	1717	2318	3436	3705	4000	4304	4667	4821	5232	5495	6153	6507	
				his-3															***cog***		

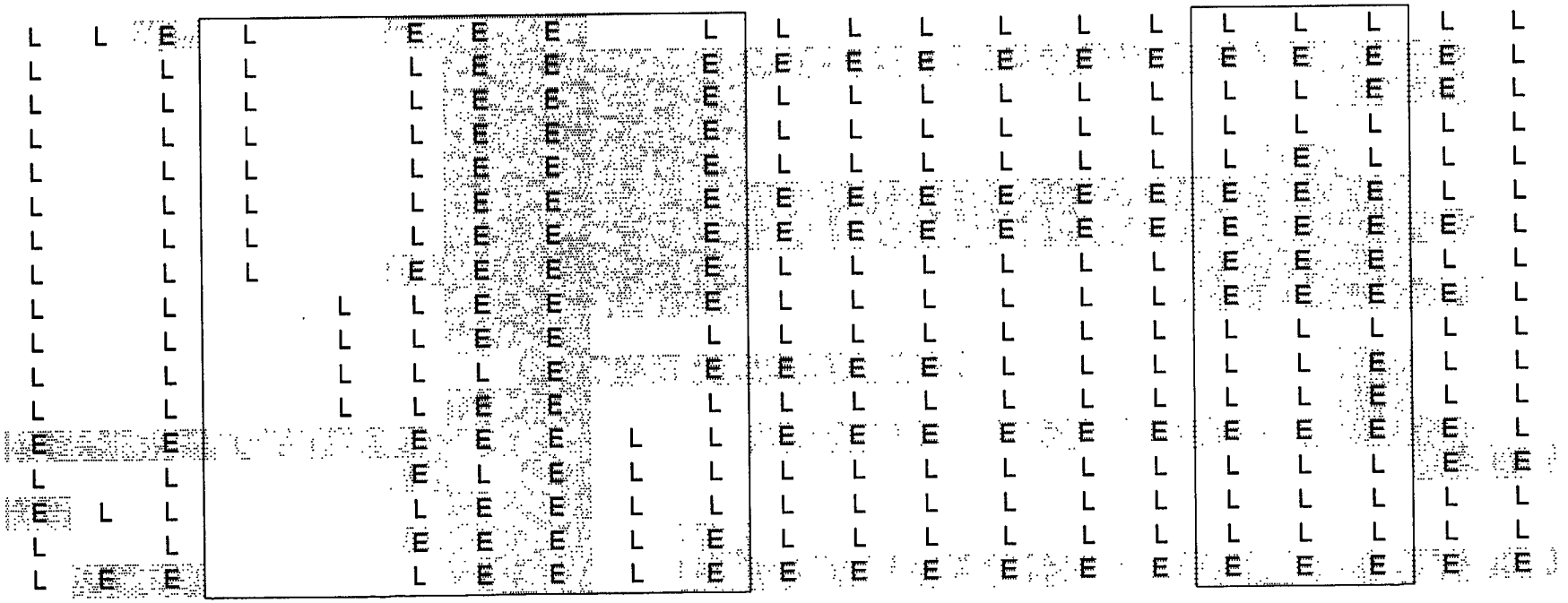


FIG. 7

Nucleotide sequence of the *his-3 cog^L lpl* region of linkage group I in the Lindegren wild type strain of *Neurospora crassa*. This differs from that in the StLawrence strain in many positions, summarised in figure 5. The coordinates of relevant features are given in the text. This sequence contains the high frequency recombinator *cog^L* which is active providing the cross in which meiosis occurs is homozygous*rec-2*.

```
   1 GATCGCAACT GGAGATCACT CGCACCGTGC CGCAGAACAA GGGCGACGAG CCTCAGGGCA
  61 GTTTAGCCTG CCGTAACAGC ACAGACCATA GCTTATTTTC ACCTGGGCGG GCGGGCGACG
 121 GCGGCACTGA CATCGGCAAG GCGGCATCAA GCAACCCCTC TGTTGCTTGC CAGCTGCCGG
 181 CCAACGTCAG CGGTACAAGG AGAAATCTGG AAGGAAAGAC TTCTGGCACC GACAGGATGG
 241 CACGCGGGAA AAGTTCCCAA TGCATGAGAT GAGGGGCATT TGCATTGCCT CCCGTCACAC
 301 TGCCCGCGAA CCCCAACCCC ACCATAGCGT CTGTCGATAC ATGGAGCGCG AAGTCGAGAA
 361 ACCTGTAATT CCTGGTAACT TTCAGGTACA CAGTACGTAC TGATCCTGGT ATCAAACCTT
 421 GCCTGCCGAG TTTTCGACGG AAAGAGGTGT GAATTGTGAA AGAGTCATAC CAAATCACCC
 481 GATTTTCATA AAGCCCGAGT CTTTTCTGTA CATAAGCGAC ACTCGAAGCG GGCCTCATCT
 541 TCATAGCCTG ATAGCTTGTA ATACTCCATC CTCGTATCTC ACTTGACCTT GAGTTCAACC
 601 CCACGTCAGA CTTCACCCGA CACATCGACG GATTGGGGAA CAGCACAATA CCTGAAAAGC
 661 GAGAAAACCA AACAGAGGAA AACACCATGG AGACAACACT TCCCCTCCCC TTCCTCGTCG
 721 GTGTCAGTGT TCCTCCCGGA CTGAATGACA TCAAGGAGGG CCTCAGCCGG GAGGAAGTCT
 781 CGTGTCTTGG CTGCGTCTTC TTCGAGGTCA AGCCCAAGAC CCTTGAGAAA ATCGTGCGAT
 841 TCCTCAAGCG TCACAATGTC GAATTTGAGC CCTACTTCGA TGTAACAGCC CTCGAGTCTA
 901 TCGATGATAT TATCACTCTT CTGGACGCCG GCGCCCGCAA GGTGTTTGTC AAGACCGAGC
 961 AGTTGGCCGA CCTCTCCGCA TATGGCTCCC GCGTTGCCCC CATTGTCACT GGAAGCAGCG
1021 CTGCTTTGCT TTCCTCCGCC ACCGAGAGCG GCCTTTTGCT CTCCGGCTTC GATCAGACTG
1081 CCTCCGAGGC TGCACAGTTT CTGGAGGAGG CCAGAGACAA GAAAATTACC CCCTTCTTCA
1141 TCAAGCCCGT TCCTGGGGCC GATCTCGAAC AGTTCATCCA GGTCGCCGCC AAGGCTAACG
1201 CCATCCCCAT CCTGCCATCC ACTGGCTTGA CAACAAAGAA GGACGAGGCC GGAAAGCTTG
1261 CCATCTCCAC CATCCTCTCG AGCGTCTGGA AGTCTGACCG TCCCGATGGT CTGCTCCCCA
1321 CCGTTGTCGT TGATGAGCAC GACACTGCTC TGGGTCTGGT CTACAGCAGT GCCGAGAGTG
1381 TGAACGAGGC CCTCAGGACA CAGACTGGTG TCTATCAGAG CCGGAAGCGC GGTCTCTGGT
1441 ACAAGGGTGC TACTTCCGGA GACACTCAGG AGCTCGTCCG CATCTCGCTT GACTGCGATA
1501 ACGATGCTCT CAAGTTTGTC GTGAAGCAGA AGGGTCGTTT CTGCCACCTC GATCAGTCCG
1561 GCTGCTTTGG TCAGCTCAAA GGCCTTCCCA AGCTCGAGCA GACTTTGATT TCGAGGAAAC
1621 AGTCTGCCCC CGAGGGCTCC TACACTGCCC GTCTCTTCTC CGATGAGAAG CTAGTCCGGG
1681 CCAAGATCAT GGAGGAGGCT GAGGAGCTCT GCACCGCTCA GACCCCCCAG GAAATCGCCT
1741 TTGAGGCTGC CGATCTCTTC TACTTTGCTC TTACCAGGGC CGTTGCTGCC GGCGTTACTC
1801 TTGCCGATAT CGAAAGGAGC CTTGACGCCA AGAGCTGGAA GGTCAAGCGC AGGACTGGAG
1861 ATGCTAAGGG TAAGTGGGCT GAGAAGGAGG GCATCAAGCC TGCGGCGTCC GCTCCCGCTG
1921 CCACTTCGGC CCCTGTCACC AAGGAGGCCG CCCAGGAGAC CACCCCTGAG AAGATCACCA
1981 TGAGACGTTT CGACGCCTCC AAGGTCTCTA CCGAGGAGCT CGATGCTGCT CTCAAGCGTC
2041 CTGCGCAAAA GTCGTCCGAT GCCATCTACA AGATCATTGT CCCCATCATC GAGGACGTCC
2101 GCAAGAACGG CGACAAGGCT GTTCTGTCGT ACACTCACAA GTTCGAGAAG GCTACCTCTC
2161 TTACTAGCCC CGTCCTGAAG GCGCCCTTCC CCAAGGAGCT TATGCAGCTC CCTGAGGAGA
2221 CCATTGCTGC CATCGACGTG TCCTTCGAGA ACATCCGCAA GTTCCACGCC GCCCAGAAGG
2281 AGGAGAAGCC CCTCCAGGTC GAGACCATGC CCGGTGTTGT CTGCAGCCGT TTCTCTCGTC
2341 CCATCGAGGC CGTCGGCTGC TACATCCCCG GCGGTACCGC CGTTCTCCCC AGCACTGCCC
2401 TTATGCTGGG TGTTCCCGCC ATGGTCGCCG GCTGCAACAA GATTGTGTTC GCCTCTCCTC
2461 CCCGCGCCGA CGGAACCATC ACTCCCGAGA TTGTCCACGT CGCTCACAAG GTTGGGGCCG
2521 AGTCCATCGT GCTTGCCGGC GGTGCCCAGG CCGTAGCTGC CATGGCCTAC GGCACCGAGA
2581 GCATCACCAA GGTCGACAAG ATTCTCGGCC CCGGTAACCA GTTCGTCACT GCTGCCAAGA
2641 TGTTCGTCAG CAACGACACC AACGCTGCCG TTGGGATTGA CATGCCCGCT GGCCCGTCCG
2701 AGGTGCTGGT CATCGCTGAC AAGGACGCCA ACCCCGCGTT CGTTGCCTCG GATCTCCTGT
2761 CCCAGGCTGA GCACGGCGTT GACAGTCAGG TCATCCTGAT CGCTATTAAC CTCGACGAGG
```

FIG. 7 continued

```
2821 AGCATCTTCA GGCTATTGAG GACGAGGTTC ACCGTCAGGC TATGGAGCTT CCTCGCGTCC
2881 AGATTGTCCG TGGCTCCATC GCCCACTCGA TCACCGTGCA GGTCAAGACC GTCGAGGAGG
2941 CCATGGAGCT CAGCAACAAG TACGCTCCTG AGCACTTGAT CCTCCAGATC AAGGAGGCCG
3001 AGAAAGCTGT CGATCTTGTC ATGAACGCTG GTAGTGTCTT CATTGGCGCT TGGACTCCTG
3061 AGTCCGTTGG CGATTACTCT GCTGGTGTTA ACCACTCGCT GCGTAAGTTA CATATCATAA
3121 ATAGCCCCGC TTCACAGATT CTTCTGCTAA CGTCAAGACA CATAGCTACC TATGGTTTTG
3181 GCAAGCAGTA CTCTGGCGTC AATCTCGCCT CGTTCGTCAA GCACATTACC AGCTCCAACT
3241 TGACTGCCGA GGGTCTCAAA AACGTCGGCC AGGCTGTCAT GCAGTTGGCT AAGGTTGAGG
3301 AGCTCGAGGC TCACAGAAGG GCGGTCAGCA TCCGTCTTGA GCACATGAGC AAGAGCAACT
3361 AGACGGAAAT TCTTTTTCGA AGTTGCAAAA AAAACAAGAA CAAAAGGATG TAGTGGGTTG
3421 ATGTATATCT GGGTCATTTT GGGCACATAG AGTAATGATA ACGAGTTTTG GACATTGTAC
3481 TGTTCTGTAC AGGCTGAAGA TCAGTACATG AATCTGTTGG TAAGTGTAGA GACCCAAACG
3541 TCCCTTGAGT TTTTCTCCCT GTTCCAGAGA GGTGCTCGTC CCTGGGTGTT TATTTTCATT
3601 ATTACATCAA CCTTTTATTT TATTTTATTT TTTATTTTAC TTTTTTTTCC TTTTTTTCAG
3661 ATCATGCGTA CATGAACGGG GGAAGCACAG ACGATCGAAA CGTGGATGTC ACAATGTCGC
3721 TGCAGTGATG CTGCATTGCA TGAAGCGCCC ATCTCAATAT ACTTGCAGTC TTGCGCGTTG
3781 CACGTGAACT TCCCAAACAA CCGAATAAAA GACGGCGAAA AATGAAGATA AAAAAAAACC
3841 ATAATAAAAA TCGGAGGGAG TGTGGGAAAT GGTTTCTTTT AGCATTTAGA CCCCATAGCC
3901 GTGCACGCCC GGGTACAGAC AGGTTCATCG ATGTTGACAT TGACTGGGAC ACCAGGTCTA
3961 TCTATTTCAT CTCCTGTCCT CTACCATACA TCGGGACATC GGACATCTCG CTGTACCCCC
4021 CACACCCACA AAGTCTTATA AAAGCGCCAC ACCCGAGGAG GTTCGGTCGG CCCCACGAAC
4081 TCCGTGCCTC CCTGCCTGTT TACAGGGACC GAACGCTGGA GAAGCTTAGT TTCCTGACAT
4141 CCGGCCTACC CGAGCAGGAA AAGGGACAGC TCATAGGCGA GGAGGGATTT GAAGATGGGG
4201 ACATTTTGGA TGATTCGAGA GGAGGAACTA GGTACTGTAT CATGATAGTT CGGGGCAGCA
4261 TCTTGGCTGG GACATTGTTA ATACCTCGAT ATGATGAAGT GGGAGGGAGT TTTTTCATGT
4321 CTTGCCCAAG TCCCACTAAT CTTTTTTTTT TTTTGTACCA ACACCCAAGA TTCGGAGAAT
4381 AGTGTAAGGA TTCGCATTCA CAAGTGGAAG TCTGAGGATC TTTTTATATC TTTGTCTTCC
4441 GCGGACTGTT AACGATCCTA CAGCGAGCGA GCGAGCGGTC GGATGCGCTG ATCTGATAGG
4501 TGCAATATAC GGCCGCTTTC TCCGGTCGTG TAGTGTAAGC TCTGTCGGCA TAGTAGTACA
4561 CTAAAAAAAC CCTTGCATTT CATGATCTGC TTGCTATTCA TTCCGAGTTA TTTCAGTGGT
4621 CACATTTCGA GATTCACAGC CATCCATCCA TATGGAAAAA TCCATTCCCA TGCTTCCTCC
4681 CCCCCACTAT GTATGTGACC ACACGCTGCT GTCAGAATGC CAACGGTCTC AGGTACCCTC
4741 GTCCGACTGT TTGGCATGGA GTTACATACA CTACTAGTGT AGCCCCGGGC CAAGCTACCC
4801 CGTCAAATCT ATACATATCT ATAATGGGTT TCAGGTGTTT CGTTCGCTGT CAATCAAGTT
4861 TGAAACATCA CTGGGGCCGT TGGACGGTGT ATTAGACCAT TGGCTCCCTC AGCTGGCGGC
4921 TGGGCGGTTG GGTCGGCAAT AACGGGACTG GACTTGAGAG GGACGAGGAG AGTCGGTTGG
4981 CTGCCTACAC TACACTACAA GCGTTCCCAC CTAACCGACG AGTCCCGTTT TCCATTTGTG
5041 TGCCTTAACC ATCATCTAGG GATGTCAGGG TTTGGCCGGA TCAGGGTATG TTTGGTTGAC
5101 TGTTGTCATG TCTGATTGGG TACATATCAT GGTAGGTGTC TCGAGAACAG TAGAGTACTC
5161 GGGCCTAGCG TTTGGATGAT TACGCGAGAT ATGAGTTGTA GGCCGCCATG CAGTTGCTTG
5221 CCCATAAGCA GAAGTTGCTT TGGGATATAT TTCTCGTCTT TCAAAGGTCA CGAGGTCCTG
5281 GGACGAGCGG CATCGCCATC CAAAGGGTTG AACATGAGAA ACCGGAATGG CCTTTGCGTT
5341 GAAATACAAA AAGTCAAGAA TAAAATCGCT TGAGGATAGG GACGTGGAAG CAAGCAAATA
5401 TGGTAAGGGA GGTACTGCTA TGTAGGTGCT CAGCAAACTG CCAATTTCTT GGCCCCCAAG
5461 CAGCAGTTTG CTGTCAGTGC TGCTCGTGTC AGCCTTGGTA GTGGAACCTA AACTGCTAAC
5521 ACAGCGCAAG TGCGCATGTA AAGATATTGT GGGAGGATCT GTATGGATGG ATGAGATTAC
5581 TGCTTGGTGT TGGTTGCGAG GCACTGCGGC TGTTAGGCTT TGCTGTGCCC CGTTCGACGA
5641 AGAAATACGC GGAACTATAA ATTGGATACC TAGACTTACT GCCTATGGGA GGTATCTACC
5701 GACGTAGCCG ACGGATTCTA GCAACATCCC GACTTTGCTT GTAGTGTACT ATGATAGCAG
5761 CACAGTGGGG TGTTGCTCCT TGTGAGCATG GGCTCTTTTT TTTTTTTTCC CCCTTCCCTA
5821 GGGCGTTGAC TGGACTTGCT CTATCGTTCC CAAGGTAGGT GCCCGTCATC GATTTTCCCA
5881 AGCCGTCTCC CGCCAGATTG TCGTCATAGT GTCATGATGA CCTCGGTCGC TGGGGCTGCG
5941 TGGTTACGGG GAGCTGGGAC CGCTAGGCCT CAGTGGTTGT GCCATTCAGC GTGGGTGTGT
6001 GGAGTAGCGG TAGAGGCGCT TGGAAGTTGT GCTAGCGGAA ACCCTGGAAT ATCTTGTACC
```

Sequence Listing

FIG. 7 continued

```
6061 CTTCGATTCC TTCTCGGGCT GCCCATGTGC TGAGGTGATG CCGGGGATCT GGCGCCAATC
6121 ATCCATTGAG GTTCCCGCAG CTTCCCGGTG CCGCGCGCGG GCGCAGTTGC TCACAGGACA
6181 CACCTAGACG CAGGGGCACA GGGGCACCGT TTGGTGTGCA ACTGGGTACC TGGTAGCTGT
6241 AGCAAGCACT CCACCGTCTG TGCAATCCCC CAATCCACGG CAGGAACTTA GCACCGCCGC
6301 GGCACCGAGT GAGCGAATCC ATCCGCATTG GATCCCAATT CTTGCCCTTG CCATCCTTCT
6361 TTCTTCCCAC TTGGCGCAAC CAACACTTCC CTTGGTCTGG GTACTCGTGT TGATCTTCAC
6421 TCTCTTTTTT TCTTGGGCGA CCGACTTTTT ATATCCGTCC TTGCTTCCCC CTGGCCGTTG
6481 TCGTTCTTTC TACAACTACC TTCCGTTCAT TATCCCCTTT CTTGGTTCGG TCGAGGACCC
6541 AAAAACAGAA CAATTCCGGC TCTTCCAGGT GGCTTGGGTG CGACTGTTTA GCTCTTGACC
6601 ACTAGCCGCT TACCTTCTCT TGATGTTTAT ATTTGGATAT CATTGAACTA CTCTTTCTTG
6661 AAACGGCAGA CGAACGGAAC AGTCCCTACG GTTTATTAGC GATATACGTT GTACTGATAT
6721 CCTGAGCAAG AAGAGGCAAA TTATCAATTA TGCATCTCCC ATCGTCGCTG CTCATCGCAG
6781 CTCCCTTGCT CGCCAATGTA TCGGCCGAAC CGATTAGGAT ACCCCAACGC GATGTTCTCC
6841 GTGGTATCAA CATCACAGCA ACTTGCCGTT CGAGCACTAC CGAATTCGCC CAGCGGTGGA
6901 TATGCCCCTG CCGTTGTAGA CTGTCCCAAG ACCAAGCCGA CGCTCCGGAA GGCCGTGGAT
6961 TTGTCGAACG AGGAGAAGAA CTGGTTGTCG ATCCGGAGGA AGAACACCAT CCAGCCCATG
7021 AGGGACCTAC TGAAGAGGGC CAACATCACT GGGTTCGATT CCGAAACTTT CATGAATGAG
7081 GCCGCCAACA ACGTCTCGCA ACTGCCCAAT GTCGCCATTG CCATTTCAGG AGGCGGCTAT
7141 CGTGCCCTCA TGAACGGCGC CGGCTTCGTT GCTGCTGCGG ATAACCGGAT TCAAAATACC
7201 ACGGGCGCAG GTGGTATTGG AGGCTTGTTG CAGTCCAGCA CATATTTGTA TGTAAAACCA
7261 TGCCTTCTTG TGGTTCTTCT TATCTCGTTT TCGAGTGTCA ACTGCGCCAG TTCGACGTTG
7321 GGCGGCTGTG GACGACCTTG CTGGTGAACA TGTCTTGGAC TCCATGCCCC TTTTTTCGTT
7381 CCCTAAAATC CCAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAATTCGAG
7441 GACCGTGACT GTAAATTGCT AACGCAACTC TAGGGCCGGA CTTTCTGGTG GTGGCTGGCT
7501 TGTCGGCAGT TTGTTCTCCA ACAACTTCAG TAGCATTGAG ACCCTGCTGA GCGAGAACAA
7561 AGTCTGGGAC TTTGAGAACT CCATCTTTAA AGGACCCAAG GAGGCTGGCC TTAGTACTGT
7621 CAACCGTATC CAGTACTGGT CCGAAGTGGC AAAGGAAGTT GCGAAGAAGA AGGATGCTGG
7681 CTTCGAGACA AGTATAACAG ACTACTGGGG CCGAGCATTG AGTTACCAAC TGATCGGAGC
7741 CGATATGGGC GGCCCGGCTT ACACCTTCTC CAGCATTGCC CAGACCGACA ACTTCCAGAA
7801 GGCCGAAACG CCGTTCCCTA TTCTGGTAGC TGACGGCCGC GCGCCTGGAG ACACCATCAT
7861 CTCCCTCAAT GCTACCAACT ACGAGTTCAA CCCGTTCGAG ACGGGTAGCT GGGACCCGAC
7921 CGTCTATGGC TTTGCGCCGA CCAAGTACCT CGGCGCCAAC TTCAGCAACG GCGTGATCCC
7981 ATCGGGAGGC AAGTGCGTTG AGGGTCTCGA CCAAGCCGGC TTCGTCATGG GCACCAGCAG
8041 CACGCTCTTC AACCAGTTCC TTTTGGCCAA CATCTCCAGC TACGACGGTG TTGCCAGACG
8101 TGCTCATCGA GGCCGTGACT TCTGTCCTCA AGGAAATCGG CGCCAAGAGG ACGACGTCTC
8161 CCAAATCATC CCTAATCCGT TCCTGGACTG GAACAACCGG ACCAACCCCA ACGCCGACAC
8221 GCTCGAGCTC GACCTGGTCG ACGGCGGCGA AGATCTGCAG AATATTCCGC TCAACCCGCT
8281 CACCCAACCC GTGCGCGCCG TCGACGTCAT CTTCGCTGTC GACTCGTCCG CCGACGTGAC
8341 AAACTGGCCC AATGGCACCG CCCTGCGCGC CACCTACGAG CGCACTTTCG GCTCTATTTC
8401 CAACGGGACA CTCTTCCCCT CGATCCCCGA CGACTGGACG TTTATAAACC TAGGCCTCAA
8461 CAACCGCCCC TCTTTCTTCG GCTGCGATGT TAAGAACTTT ACCTTGAACG CCAACCAAAA
8521 GGTTCCCCCC TTAATCGTCT ATGTCCCCAA CGCGCCCTAT ACCGCGCTGA GCAACGTGTC
8581 CACCTTCGAT CCGTCATACA CGATGTCTCA GCGCAACGAC ATCATCGGCA ACGGATGGAA
8641 CTCAGCCACG CAGGGAAACG GCACGCTGGA TTCGGAGTGG CCCACTTGCG TCGCCTGCGC
8701 GGTTATCAGC AGGAGCTTAG ATCGGTTGGG CAGGCAGACG CCAGCCGCGT GCAAGACTTG
8761 CTTTGACAGG TATTGCTGGA ATGGCACAGT GAACTCCAAA GATACGGGGG TTTACATGCC
8821 TGAGTTCAAG ATTGCGGATG CGCATGCCCT GGACTCGGGT GCTGTTGCTA TCGGAAAGAT
8881 GGTGAATGTC TGGTCGTCGG TTGTGGTGGG AGTTGTGGCG GCTACTTTGT TGTTGTAGGG
8941 GTAGGGGAGA CGTGATGATA TTCCAGTCTG ATGAAGTTGA GACTGGACTG GAGATCGCCA
9001 AGGATGCGGA GGGAAAGGAA TGCGTGGTGT TAATGTCATG ATGGATGAAG AGTCATGGAT
9061 CATGGAACGA CGGGGCGGGG ATATTGGATG ATGGATATAC CACACTGCAT GCATGCTCTA
9121 TTGATAGTAT GCTTTGGCAT TTACGTTTAA CAATCAATTG CTCCATCCTG ATGTTCTATC
9181 TTTTTCGACA ATGGATTGAT ACTACTCCTG TTGCTTCGCT CTTGAGGTTG GAAGGACTTG
9241 AGGTTGGAAG GACTTGAGGT TGTTTGTTCT GAGGGAGGTT ATCGAAGTAT CATCTGTGCT
```

FIG. 7 continued

```
9301 GATGCCGATT GATAGACTGT CCTCTTCTTC GAGGCAACGA ACGGTCGGAT GAGCCTCTTT
9361 AATCATGATG CTCAGTGCCA CAAAAAGGCT CCAGCACAGC TGCCCACACC TTTCTTGCCT
9421 CGCCGTTCCT TCCTTTTTCT TTTCCCCTGT TTCCTTTCTT CCTTTCCATC TCATCCCGTA
9481 CCAGAGTGCC CACCGGGTAT ATATATTACC TCCTTGGCCG TTCTCCTTTG ACCAATAAAT
9541 CGCTTGGTCG AGTGGCGTAA CGGTTTACCG TCTACACTTA TCACTCAAAC CAAACCAAAC
9601 CATCGAAGAA GTGACCTATC GGTTCGAGGG AACGGTGATG TTCTTACGAC CAAGTTAACC
9661 CAAAGAGCGT TCCACATCGT TGAACCGTCT CCTCCAGTTG GATCTGTTTA ACTTCCGCAG
9721 CGACTGAAGA AGGTATCACT TTTTTTTTGG TTCCAAAAAA AAAAAAAAAA ATTAC
```

5,565,340

FIG. 8

Nucleotide sequence of the *his-3 cog^E lpl* region of linkage group I in the StLawrence wild type strain of *Neurospora crassa.* This differs from that in the Lindegren strain in many positions, summarised in figure 5. The coordinates of relevant features are given in the text. This sequence contains the weak recombinator *cog^E* and also the remnant of a transposable element *Guest* within the replaceable sequence 3' of *his-3*. StLawrence strains carry *rec-2$^+$* which prevents the initiation of recombination at *cog*.

```
   1 ACCGGGAATC GTAGCGGGCG CTAAGGCCAA GCCGCGGCAC GGGTCACTGA CCCAATGCAG
  61 CGCATTCCGT CAGCAACTGA AGTGGATGTA CAAGTACATA GTAGTAGATC GCAACTGGAG
 121 ATCACTCGCA CCGTGCCGCA GAACAAGGGC GACGAGCCTC AGGGCAGTTT AGCCTGCCGT
 181 AACAGCACAG ACCATAGCTT ATTTTCACCT GGGCGGGCGG GCGACGGCGG CACTGACATC
 241 GGCAAGGCGG CATCAAGCAA CCCCTCTGTT GCTTGCCAGC TGCCGGCCAA CGTCAGCGGT
 301 ACAAGGAGAA ATCTGGAAGG AAAGACTTCT GGCACCGACA GGATGGCACG CGGGAAAAGT
 361 TCCCAATGCA TGAGATGAGG GGCATTTGCA TTGCCTCCCG TCACCCAGTG CGAACCCCAA
 421 CCCCACCATA GCGTCTGTCG ATACATGGAG CGCGAAGTCG AGAAACCTGT AATTCCTGGT
 481 AACTTTCAGG TACACAGTAC GTACTGATCC TGGTATCAAA CCTTGCCTGC CGAGTTTTCG
 541 ACGGAAAGAG GTGTGAATTG TGAAAGAGTC ATACCAAATC ACCCGATTTT CATAAAGCCC
 601 GAGTCTTTTC TGTACATAAG CGACACTCGA AGCGGGCCTC ATCTTCATAG CCTGATAGCT
 661 TGTAATACTC CATCCTCGTA TCTCACTTGA CCTTGAGTTC AACCCCACGT CAAACTTCAC
 721 CCGACACATC GACGGATTGG GGAACAGCAC AATACCTGAA AAGCGAGAAA ACCAAACAGA
 781 GGAAAACACC ATGGAGACAA CACTTCCCCT CCCCTTCCTC GTCGGTGTCA GTGTTCCTCC
 841 CGGACTGAAT GACATCAAGG AGGGCCTCAG CCGGGAGGAA GTCTCGTGTC TTGGCTGCGT
 901 CTTCTTCGAG GTCAAGCCCC AGACCCTTGA GAAAATCCTG CGATTCCTCA AGCGTCACAA
 961 TGTCGAATTT GAGCCCTACT TCGATGTAAC AGCCCTCGAG TCTATCGATG ATATTATCAC
1021 TCTTCTGGAC GCCGGCGCCC GCAAGGTGTT TGTCAAGACC GAGCAGTTGG CCGACCTCTC
1081 CGCATATGGC TCCCGCGTTG CCCCCATTGT CACTGGAAGC AGCGCTGCTT TGCTTTCCTC
1141 CGCCACCGAG AGCGGCCTTT TGCTCTCCGG CTTCGATCAG ACTGCCTCCG AGGCTGCACA
1201 GTTTCTGGAG GAGGCCAGAG ACAAGAAAAT TACCCCCTTC TTCATCAAGC CCGTTCCTGG
1261 GGCCGATCTC GAACAGTTCA TCCAGGTCGC CGCCAAGGCT AACGCCATCC CCATCCTGCC
1321 ATCCACTGGC TTGACAACAA AGAAGGACGA GGCCGGCAAG CTTGCCATCT CCACCATCCT
1381 CTCGAGCGTC TGGAAGTCTG ACCGTCCCGA TGGTCTTCTC CCCACCGTTG TCGTTGATGA
1441 GCACGACACT GCTCTGGGTC TGGTCTACAG CAGTGCCGAG AGTGTGAACG AGGCCCTCAG
1501 GACACAGACT GGTGTCTATC AGAGCCGGAA GCGCGGTCTC TGGTACAAGG GTGCTACTTC
1561 CGGAGACACT CAGGAGCTCG TCCGCATCTC GCTTGACTGC GATAACGATG CTCTCAAGTT
1621 TGTCGTGAAG CAGAAGGGTC GTTTCTGCCA CCTCGATCAG TCCGGCTGCT TTGGTCAGCT
1681 CAAAGGCCTT CCCAAGCTCG AGCAGACTTT GATTTCGAGG AAACAGTCTG CCCCCGAGGG
1741 CTCCTACACT GCCCGTCTCT TCTCCGATGA GAAGCTAGTC CGGGCCAAGA TCATGGAGGA
1801 GGCTGAGGAG CTCTGCACCG CTCAGACCCC CCAGGAAATC GCCTTTGAGG CTGCCGATCT
1861 CTTCTACTTT GCTCTTACCA GGGCCGTTGC TGCCGGCGTT ACTCTTGCCG ATATCGAAAG
1921 GAGCCTTGAC GCCAAGAGCT GGAAGGTCAA GCGCAGGACT GGAGATGCTA AGGGTAAGTG
1981 GGCTGAGAAG GAGGGCATCA AGCCTGCGGC GTCCGCTCTC GCTGCCACTT CGGCCCCTGT
2041 CACCAAGGAG GCCGCCCAGG AGACCACCCC TGAGAAGATC ACCATGAGAC GTTTCGACGC
2101 CTCCAAGGTC TCTACCGAGG AGCTCGATGC TGCTCTCAAG CGTCCTGCGC AAAAGTCGTC
2161 CGATGCCATC TACAAGATCA TTGTCCCCAT CATCGAGGAC GTCCGCAAGA ACGGCGACAA
2221 GGCTGTTCTG TCGTACACTC ACAAGTTCGA GAAGGCTACC TCTCTTACTA GCCCCGTCCT
2281 GAAGGCGCCC TTCCCCAAGG AGCTTATGCA GCTCCCTGAG GAGACCATTG CTGCCATCGA
2341 CGTGTCCTTC GAGAACATCC GCAAGTTCCA CGCCGCCCAG AAGGAGGAGA AGCCCCTCCA
2401 GGTCGAGACC ATGCCCGGTG TTGTCTGCAG CCGTTTCTCT CGTCCCATCG AGGCCGTCGG
2461 CTGCTACATC CCCGGCGGTA CCGCCGTTCT CCCCAGCACT GCCCTTATGC TGGGTGTTCC
2521 CGCCATGGTC GCCGGCTGCA ACAAGATTGT GTTCGCCTCT CCTCCCCGCG CCGACGGAAC
2581 CATCACTCCC GAGATTGTCC ACGTCGCTCA CAAGGTTGGG GCCGAGTCCA TCGTGCTTGC
2641 CGGCGGTGCC CAGGCCGTAG CTGCCATGGC CTACGGCACC GAGAGCATCA CCAAGGTCGA
2701 CAAGATTCTC GGCCCCGGTA ACCAGTTCGT CACTGCTGCC AAGATGTTCG TCAGCAACGA
```

FIG. 8 continued

```
2761 CACCAACGCT GCCGTTGGTA TTGACATGCC CGCTGGCCCG TCCGAGGTGC TGGTCATCGC
2821 TGACAAGGAC GCCAACCCCG CGTTCGTTGC CTCGGATCTC CTGTCCCAGG CTGAGCACGG
2881 CGTTGACAGT CAGGTCATCC TGATCGCTAT TGACCTCGAC GAGGAGCATC TTCAGGCTAT
2941 TGAGGACGAG GTTCACCGTC AGGCTACGGA GCTTCCTCGC GTCCAGATTG TCCGTGGCTC
3001 CATCGCCCAC TCGATCACCG TGCAGGTCAA GACCGTCGAG GAGGCCATGG AGCTCAGCAA
3061 CAAGTACGCT CCTGAGCACT TGATCCTCCA GATCAAGGAG GCCGAGAAGG CTGTCGATCT
3121 TGTCATGAAC GCCGGTAGTG TCTTCATTGG CGCCTGGACT CCTGAGTCCG TTGGCGATTA
3181 CTCTGCTGGT GTTAACCACT CGCTGCGTAA GTTACATATC ATAAATAGCC CCGCTTCACA
3241 GATTCTTCTG CTAACGTCAA GACACATAGC TACCTATGGC TTTGGCAAGC AGTACTCTGG
3301 CGTCAATTTC GCCTCGTTCG TCAAGCACAT TACCAGCTCC AACTTGACTG CCGAGGGTCT
3361 CAAAAACGTC GGCCAGGCTG TCATGCAGTT GGCTAAGGTT GAGGAGCTCG AGGCTCACAG
3421 AAGGGCGGTC AGCATCCGTC TTGAGCACAT GAGCAAGAGC AACTAAACGG AAATTCTTTT
3481 CGAAGTAGCA AAAAAAAAAA AAAAAAACAA GAACAAAAGG ATGTAGTGGG TTGATGTATA
3541 TCTGGGTCAT TTTGGGCACA TAGAGTAATG ATAACGAGTT TTGGACATTG TACTGTTCTG
3601 TACAGGCTGA AGATCAGTAC ATGAATCTGT TGGTAAGTGT GGAGACCCAA ACGTCCCTTG
3661 AGTTTTTCTC CCTATTCCAG AGGTGCTCGT CCCTGGGTGT TTATTTTCAT TATTACATCA
3721 ACCTTTTTTT TTTTTTTTTT TTTTTCAGAT CATGCGTACA TGAACGGGGG AAGCACAGAC
3781 GATCGAAACG TGGATGTCAC AATGTCGCTG CAGTGATGCT GCATTGCATG AAGCGCCCAT
3841 CTCAATATAC TTGCAGTCTT GCACGTTGCA TGTGAACTTC CCAAACAACC GAATAAAAGA
3901 CGGCGAAAAA TGAAGATAAA AAAAAACCAT AAAAAAAATC AGAGGGAGTG TGGGAAATGG
3961 TGTCTTTTAG CATTCAGACC CCATAGCCGT GCACGCCCGG GTACAGACAG GTTCATCGAT
4021 GTTGACATTG ACTGGGACAC CAGGTCTATC TATTTTATCT CCTGTCCTCT ACCATACATC
4081 GGGACATCGG ACATCTTGCT GTACCCCCCA CACCCACAAA GCCTTATAAA AGCGCCACAC
4141 CCGAGGAGGT TCGGTCGGCC CCACGAACTC TGTGCCTCCC TGCCTGTTTA CAGGGACCGA
4201 ACGCTGGAGA ATCTTACTAG TTTCCTGACA TCCGGCCTAC CCGAGCAGGA AAAGGGACAG
4261 CTCATAGGCG AGGAGGGATT TGAAGATGGG AACATTTTGG GTGATTCGAG AGGAGGAACT
4321 AGGTACTGCA TCATGATAGT TCGGGGCAGC ATCTTGGCTG GGACATTGTT AATACCTCGA
4381 TATGATGAAG TAGGAGGGAG TTTTTGCGTG TCTTGCCGAA GTCCAGAGAT CTGTTTTATT
4441 TTATTTTTTA TGGATGTAGT GTATCAACAC CCAAGATTCG GAGAATAGTA CTAGGATTCG
4501 CATTTACAAG TGGAAGTCTT GAGAATCGTT GTATATCCTT GTCTTCCTCG GAATGTTAAC
4561 AATCCTACAG CGAGCGAGCG AGCGGTCGGA TGCGCTGATC TGATAGGCGC AATATACGGC
4621 CGCTTTCTCC GGTCGTGTAG TGTAAGCTCT GTGGGCATAG TACACTAAAA AAACCCTTGC
4681 ATTTCATGAT CTGCCTGCTA TTCATTCCGA GCTATTTCAG TGGTCACATT TCGAGGAAGA
4741 AAGAAAGCAA CTAAGATTCA CAGCCATCCA TCCATCCATA TGGAAGAATA ATCCATTCCC
4801 ATGTTCCCTC CCCCCCACTA TGTATGTGAC CACACGCTGC TGTCAGAATG CCAACGGTCT
4861 CAGGTACCCT CGTCCGACTG TTTGGCATGG AGTTACATAC ACTACTAGTG TAGCCCCGGG
4921 CCAAGCTACC CCGTCAAATC TATACATATC TATAACGGGT TTCAGGGGTT TCGTTCGCTG
4981 TCAATCAAGT TTGAAACATC ACTGGGGCCG TTGGACGGTG TATTAGACCA TTGGCTCCCT
5041 CAGCTGTTTG GCGGCTGGGC GGCTGGGTCA AACGGCAATA ACGGGACTCG AGAGGGACGA
5101 GGAGAGTCGG TTGGCTGGCT GCAATACAAG CGTTCCCACC TAACCAACGA GTCCCGTTTT
5161 CCATTTGTGT GCCTAACCAT CATCTAGGGA TGTCAGGGTT TGGCCGGATC AGGGTATGTT
5221 TGGTTGACTG TTGTCATGTC TGATTGGGTA CATATTATGG TAGGTGTCTC GAGAACAGTA
5281 GAGTACTCGG GCCTAGCGTT TGGATGATTA CGCGAGATAT GAGTTGTGGG CCGCCATGCA
5341 GTTGCTTGTC CATAAGCAGA AGTTGCTTTG GGATATATTT CTCGTCTTTC AAAGGTCACG
5401 AGGTCCTGGG ACGAACGGCA TCGCCATCCA AAGGGTTGAA CATGAGAAAC CTGAATGGCC
5461 TTTGCGTTGA AATACAAAAA GTCAAGAACA AAATCGCTTG AGGATAGGGA CGTGGAAGCA
5521 AGCAAATATG GTAAGAGAGG TATACATCAA CCCTGGTTCA ATTGTTAGCG TGGTTCTTCC
5581 TCCACGTCCT CGTTCATGAC GGTTAACAGT ACCAGGCTAA CAATTAAACC AGGGTTGATG
5641 TGTACTGATA TGTAGGTGCT CAGCAAACTG CCAATTTCTT TGGCCCCAAG CAGCAGTTTG
5701 CTGTCAGTGC TGCTCGTGTC AGCCTTGGTA GTGGAACCTA AACTGCTAAC ACAGCGCAAG
5761 TGCGCATGTA AAGATATTGT GGGAGGATCT GTATGGATGG ATGAGATTAC TGCTTGGTGT
5821 TGGTTGCGAG GCACTGCGGC TGTTAGGCTT TGCTGTGCCC CGTTCGACGA AGAAATACGC
5881 GGAACTATAA ATTGGATACC TAGACTTACT GCCTATGGGA GGTATCTACC GACGTAGCCG
5941 ACGGATTCTA GCAACATCCC GACTTTGCTT GTAGTGTACT ATGATAGCAG CACAGTGTTG
```

FIG. 8 continued

```
6001 CTCCTTGTGA GAATGGGCTC TTTTTTTTTT TCCCCCTTCC CTAGGGCGTT GACTGGACTT
6061 GCTCTATTGT TCCCAAGGTA GGTGCCCGTC ATCGATTTTC CCAAGTCTCC CGCCAGATTG
6121 TCGTCATAGT GTCATGATGA CCTCGGTCGC TGGGGCTGCG TGGTTACGGG GAGCTGGGAC
6181 CGCTAGGCCT CAGTGGTTGT GCCATTCAGC GTGGGTGTGT GGAGTAGCGG TAGAGGCGCT
6241 TGGAAGTTGT GCTAGCGGAA ACCCTGGAAT ATCTTCTACC CTCGATTCCT TCTCGGGCTG
6301 CCCATGTGCT GAGGTGATGC CGGGGATCTG GCGCCAATCA TCCATTGAGG TTCCCGCAGC
6361 TTCCCGGTGC CGCGCGCGGG CGCAGTTGCT CACAGGACAC ACCTAGACGC AGGGGCACAG
6421 GGGCACCGTT TGGTGTGCAA CTGGGTACCT AGCTGTAGCA AGCACTCCAC CGTCTGTGCA
6481 ATCCCCCAAT CCACGGCAGG AACTTCGCAC CGCCGCGGCA CCGAGTGAGC GAATCCATCC
6541 GCATTGGATC CCAATTCTTG CCCTTGCCAT CCTTCTTTCT TCCCACTTGG CGCAACCAAC
6601 ACTTCCCTTG GTCTGGGTAC TCGTGTTGAT CTTCACTCTC TTTTTTTCTT GGGCGACCGA
6661 CTTTTTATAT CCGTCCTTGC TTCCCCCTGG CCGTTGTCGT TCTTTCTACA ACTACCTTCC
6721 GTTCATTATC CCCTTTCTTG GTTCGGTCGA GGACCCAAAA ACAGAACAAT TCCGGCTCTT
6781 CCAGGTGGCT TGGGTGCGAC TGTTTAGCTC TTGACCACTA GCCGCTTACC TTCTCTTGAT
6841 GTTTTTATTT GGATATCATT AAACTACTCT TTCTTGAAAC GGCAGACGAA CGGAACAGTT
6901 CCTACGGTAT ATTAGCGATA TACGTTGTAC TGATATTCTG AGCAAGAAGA GGCAAATTAT
6961 CAATTATGCA TCTCCCTTCG TCGCTGCTCA TCGCAGCTCC CTTGCTCGCC AATGTATCGG
7021 CCGAACCCAT TAGGATACCC CAACGCGATG TTCTCCGTGG TATCAACATC ACAGCAACTT
7081 GCCGTTCGAG CACTACCGGA TTCGCCCAGC GGTGGATATG CCCCTGCCGT TGTAGACTGT
7141 CCCAAGACCA AGCCGACGCT CCGGAAGGCC GTGGATTTGT CGAACGAGGA GAAGAACTGG
7201 TTGTCGATCC GGAGGAAGAA CACCATCCAG CCCATGAGGG ACCTCCTGAA GAGGGCCAAC
7261 ATCACTGGGT TCGATTCCGA GACATTTATG AATGAGGCCG CCAACAACAT CTCGCAACTG
7321 CCCAATGTCG CCATTGCCAT TTCAGGAGGC GGCTATCGTG CCCTCATGAA CGGCGCCGGC
7381 TTCGTTGCTG CTGCGGATAA CCGAATTCAA AATACCACGG GCGCAGGTGG TATTGGAGGC
7441 TTGTTGCAGT CCAGCACATA TTTGTATGTA AAGTGGTTCT TCTTATCTCG TTTTCGAGTG
7501 TCAACTGCGC CAGTTCAGAG TTGGGCGGCT GTGGACGACC TTGCTGGTGA ACATGTCTTG
7561 GACTCCATGC CCCTTCTTCG TTTCCTCAAA TCAAGAAGTC GAGGACCGTG ACCGTAAATC
7621 GCTAACGCAA CTCTAGGGCC GGACTTTCTG GTGGTGGCTG GCTTGTCGGC AGTTTGTTCT
7681 CCAACAACTT CAGCAGCATT GAGACCCTGC TGAGCGAGAA CAAAGTCTGG GACTTTGAGA
7741 ACTCCATCTT TAAAGGGCCC AAGGAGGCTG GCCTTAGTAC TGTCAACCGC ATTCAGTACT
7801 GGTCCGAAGT GGCAAAGGAA GTTGCCAAGA AGAAGGATGC TGGCTTCGAG ACAAGTATAA
7861 CAGACTACTG GGGCCGAGCA TTGAGTTACC AACTGATCGG AGCCGATATG GGCGGCCCGG
7921 CTTACACCTT CTCCAGCATT GCCCAGACCG ACAACTTCCA GAAGGCCGAA ACGCCGTTCC
7981 CTATTCTGGT AGCTGACGGC CGCGCGCCTG GAGACACCAT CATCTCCCTC AATGCTACCA
8041 ACTACGAGTT CAACCCGTTC GAGACGGGTA GCTGGGACCC GACCGTCTAT GGCTTTGCGC
8101 CGACCAAGTA CCTCGGCGCC AACTTCAGCA ACGGCGTGAT CCCATCGGGA GGCAAGTGCG
8161 TTGAGGGTCT CGACCAAGCC GGCTTCGTCA TGGGCACCAG CAGCACGCTC TTCAACCAGT
8221 TCCTTTTGGC CAACATCTCC AGCTACGACG GTGTTGCCCG ACGTGCTCAT CGAAGCCGTG
8281 ACTTCTGTCC TCAAGGAAAT CGGCGCCAAG AGGACGACGT CTCCCAAATC ATCCCTAATC
8341 CGTTCCTGGA CTGGAACAAC CGGACCAACC CCAACGCCGA CACGCTCGAG CTCGACCTGG
8401 TCGACGGCGG CGAAGATCTG CAGAATATTC CGCTCAACCC GCTCACCCAA CCCGTGCGCG
8461 CCGTGGACGT CATCTTCGCT GTCGACTCGT CCGCCGACGT GACAAACTGG CCCAATGGCA
8521 CCGCCCTGCG AGCCACCTAC GAGCGCACTT TCGGCTCTAT TTCCAACGGG ACACTCTTCC
8581 CCTCGATCCC CGACGACTGG ACGTTTATAA ACCTAGGCCT CAACAACCGC CCCTCTTTCT
8641 TCGGCTGCGA TGTTAAGAAC TTTACCTTGA ACGCCAACCA AAAGGTTCCC CCCTTAATCG
8701 TCTATGTCCC CAACGCGCCC TATACCGCGC TGAGCAACGT GTCCACCTTC GATCCGTCAT
8761 ACACCATGTC TCAGCGCAAC GACATCATCG GCAACGGATG GAACTCAGCC ACGCAGGGAA
8821 ACGGCACGCT GGATTCGGAG TGGCCCACTT GCGTCGCCTG CGCGGTTATC AGCAGGAGCT
8881 TAGATCGGTT GGGCAGGCAG ACGCCAGCCG CGTGCAAGAC TTGCTTTGAG AGGTATTGCT
8941 GGAATGGCAC AGTGAACTCA AAAGATACAG GGGTTTACAT GCCTGAGTTC AAGATTGCGG
9001 ATGCGCATGC CCTGGACTCG GGTGCTGTTG CTATCGGAAA GATGGTGAAT GTCTGGTCGT
9061 CGGTTGTGGT GGGAGTTGTG GCGGCTACTT TGTTGTTGTA GGGGTAGGGG AGACGTGATG
9121 ATATTCCAGT CTGATGAAGT TGAGACTGGA CTGGAGATCG CCAAGGATGC GGAGGGAAAG
9181 GAATGCGTGG TGTTAATGTC ATGATGGATG AAGGGTCATG GATCATGGAA CGACGGGGCG
```

FIG. 8 continued

```
9241 GGGATATTGG ATGATGGATA TACCACACTG CATGCATGCT CTATTGATAA TATGCTTTGG
9301 CATTTACGTT TAACAATCAA TTGCTCCATC CTGATGTTCT ATCTTTCGAC ACTGGATTGA
9361 TACTACTCCT GTTGCTTCCC TCTTGAAGTT GGAAGGACTT GAGGTTGGAA GGACTTGAGG
9421 TTGTTTGTTC TGAGGGAGGT TATCGAAGTA TCATCTGTGC TGATGCCGAT CGATAGACTG
9481 CCCTCTTCTT CGAGGCAACG AACGGTCGGA TGAGCCTCTA ATCATGATGC TCAGTGCCAC
9541 AAAAAGGCTC CAGCACAGCT GCCCACACCT TTTTTGCCTC GTCGCTCCTT CCTTTTTTTC
9601 CCCCCCTTTC TTCCTTTCCA TCTCATCCCG TACCAGAGTG CCCACCGGGT ATATATATTA
9661 CCTCCTTGGC CGTTCTCCTT TGACCAATAA ATCGCTTGGT CGAGTGGCGT AACCGTTTAC
9721 CGTCTACACT TATCACTCAA ACCAAACCAA ACCATCGAAG AAGTTACCTA TCGGTTCGAG
9781 GGAACGGTGA TGTTCTTACG TTCAAGTTAA CCCAAAGAGC GTTCCACATC GTTGAACCGT
9841 CTCCTCCAGT TCTTGGATCT GTTTAACTTC CGCAGCGACT GAAGAAGTAA TCACTTTTTT
9901 TTTTTTGGT TCCAAAAAAA AAAAAAAAAA TTAC
```

5,317,097

FIG. 9

Construction of the components of the sequence diversification cross: Parent (variant 1) and Parent (variant 2). For convenience, plasmid sequences are shown as linear. The cross hatched region in the chromosome is dispensible. Stippled sequence in the plasmid indicate the multiple cloning site for inserting foreign DNA. Crossovers in region 1 and region 2 insert the foreign sequence to be diversified into chromosome 1 of *Neurospora crassa* adjacent to the recombination hotspot *cog* . Parent (variant 2) containing a version of the foreign sequence with multiple differences from that in parent (variant1) is simularly constructed. Parent (variant 1) and parent (variant 2) are crossed and conversion events (stippled arrow) initiated (X) at cog^L recombine the sequence differences in variant 1 and variant 2 to form new combinations. Sequences are identical except for those that distinguish variant 1 and variant 2. *rec-2* on linkage group V permits cog^L to be active. For simplicity, genes not directly related to the diversification are omitted. See text for further details.

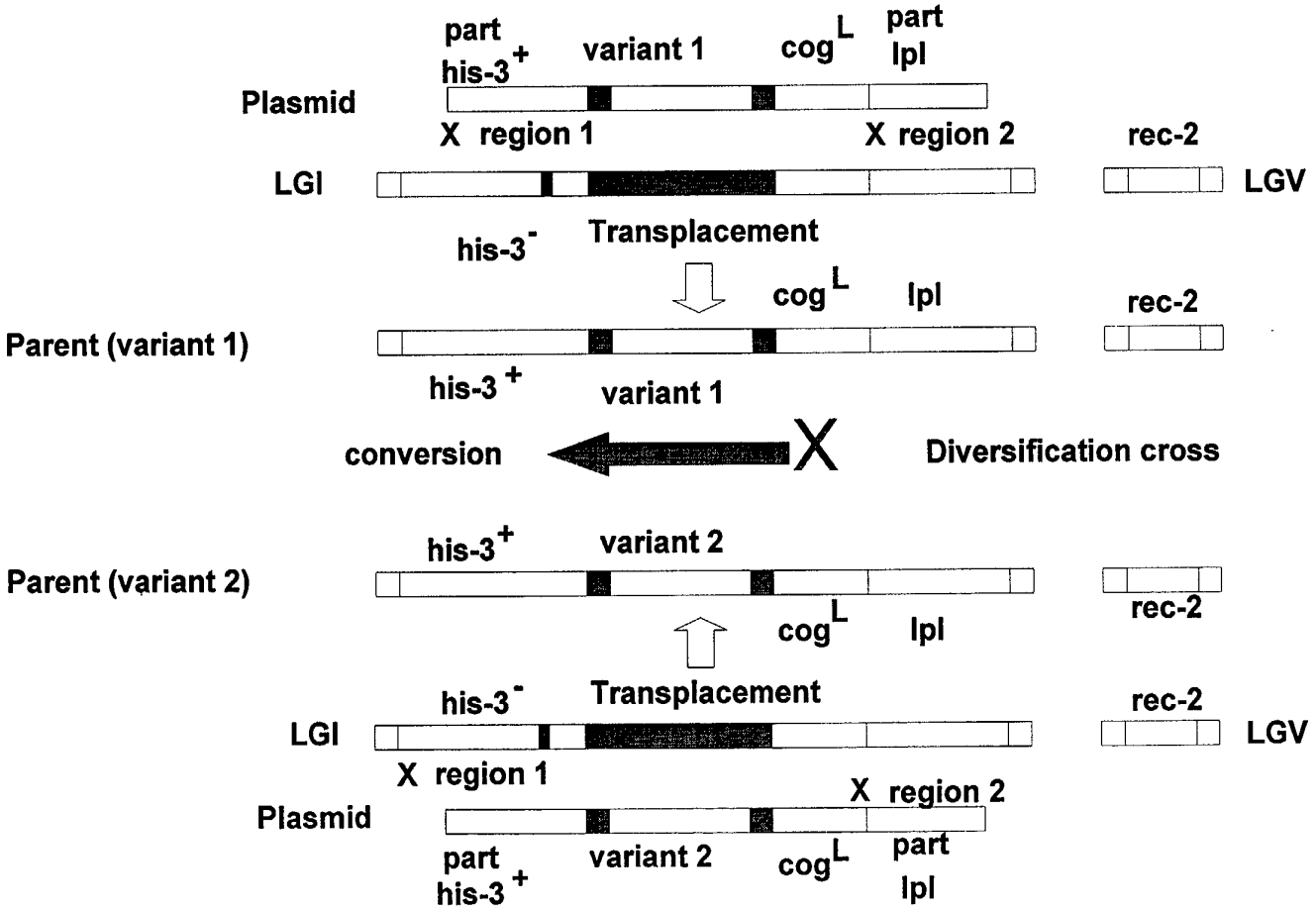
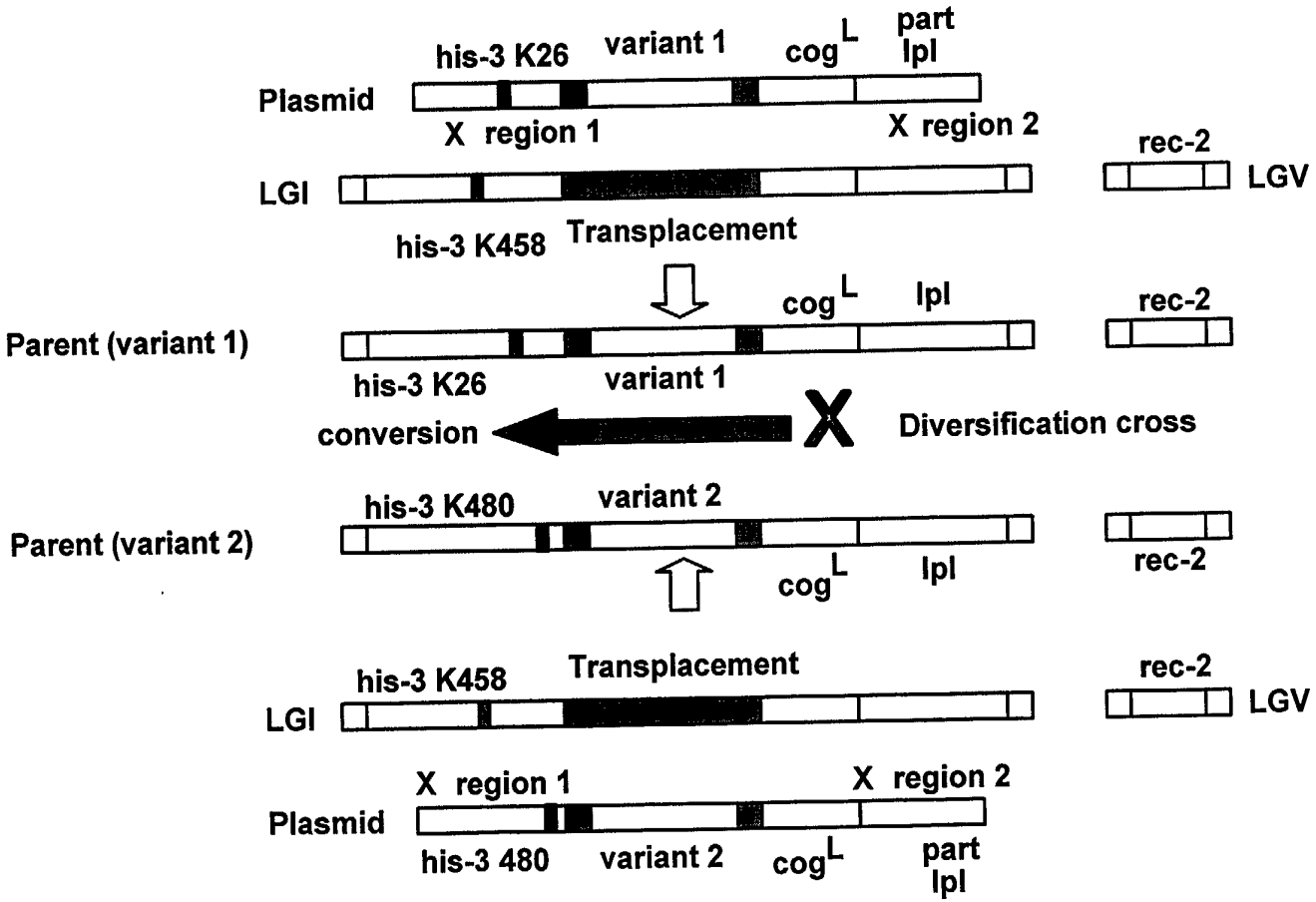


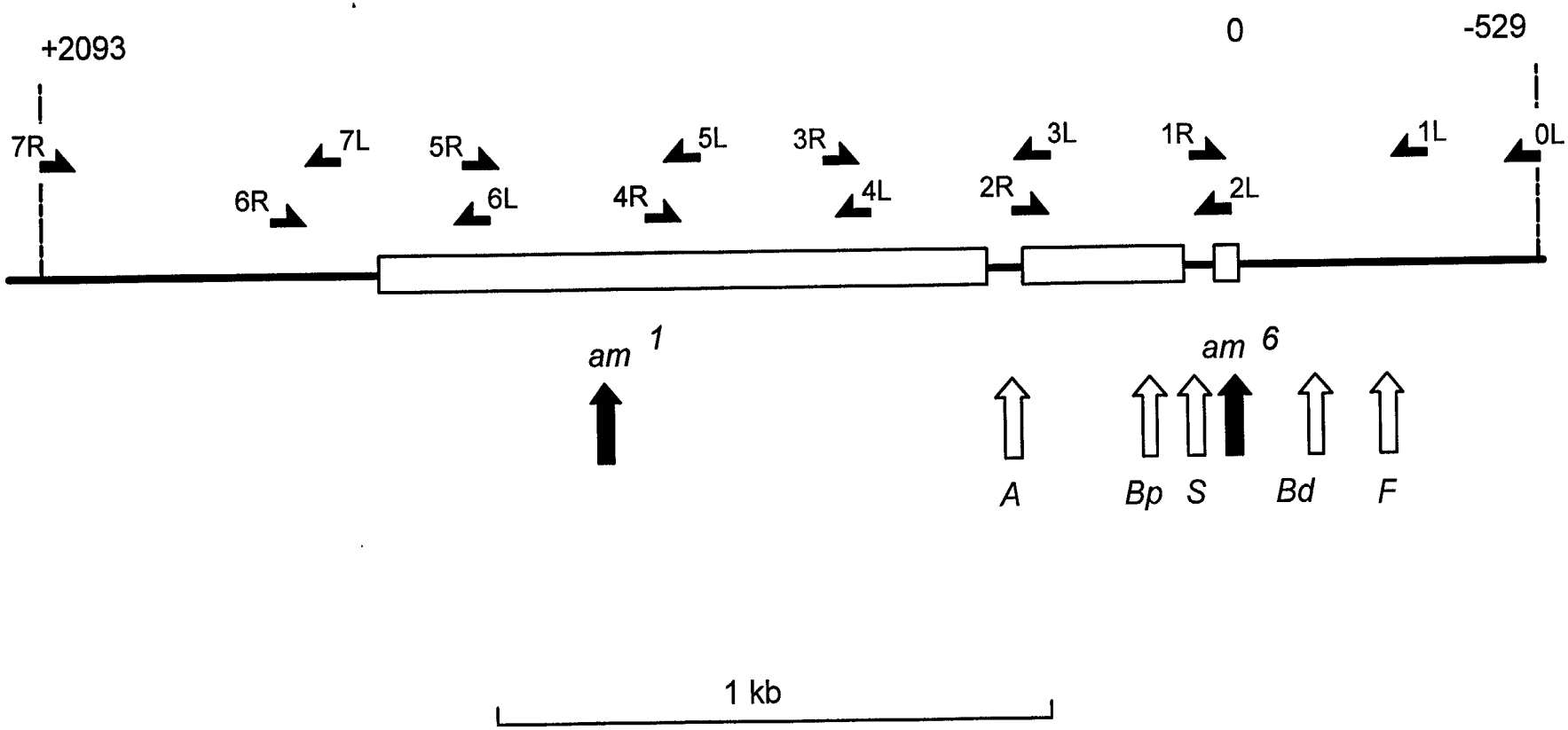
FIG. 10

Construction of parent (variant 1) and parent (variant 2) enabling selection of progeny that have experienced conversion in the foreign DNA. Complementing pairs of *his-3* alleles are used to obtain parent (variant 1) and a different pair of complementing *his-3* alleles are used to obtain parent (variant 2) as explained in the text. Parent (variant 1) and parent (variant 2) are crossed and his^+ recombinants are selected. These must all have experienced conversion events affecting the foreign DNA since the events begin at cog^L. The *his-3* alleles in parent (variant 1) and parent (variant 2) are non complementing to ensure that selection yields recombinants and not aneuploid progeny having two copies of all or part of linkage group I.



006237 3961S260

FIG. 11



006237 "3957260

FIG. 12

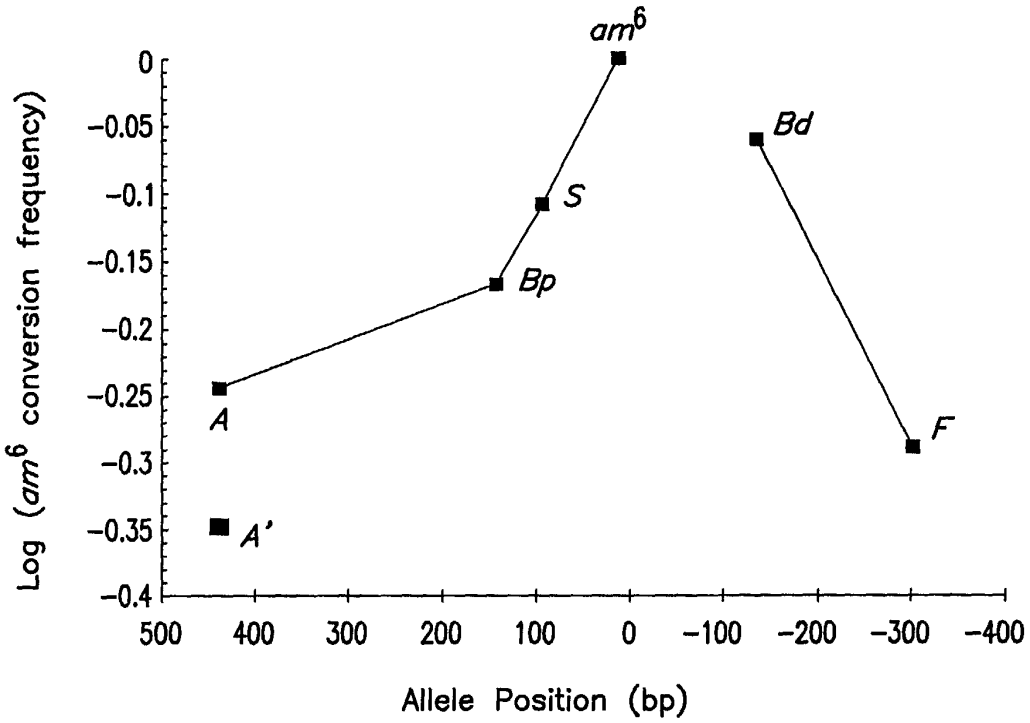


FIG. 13

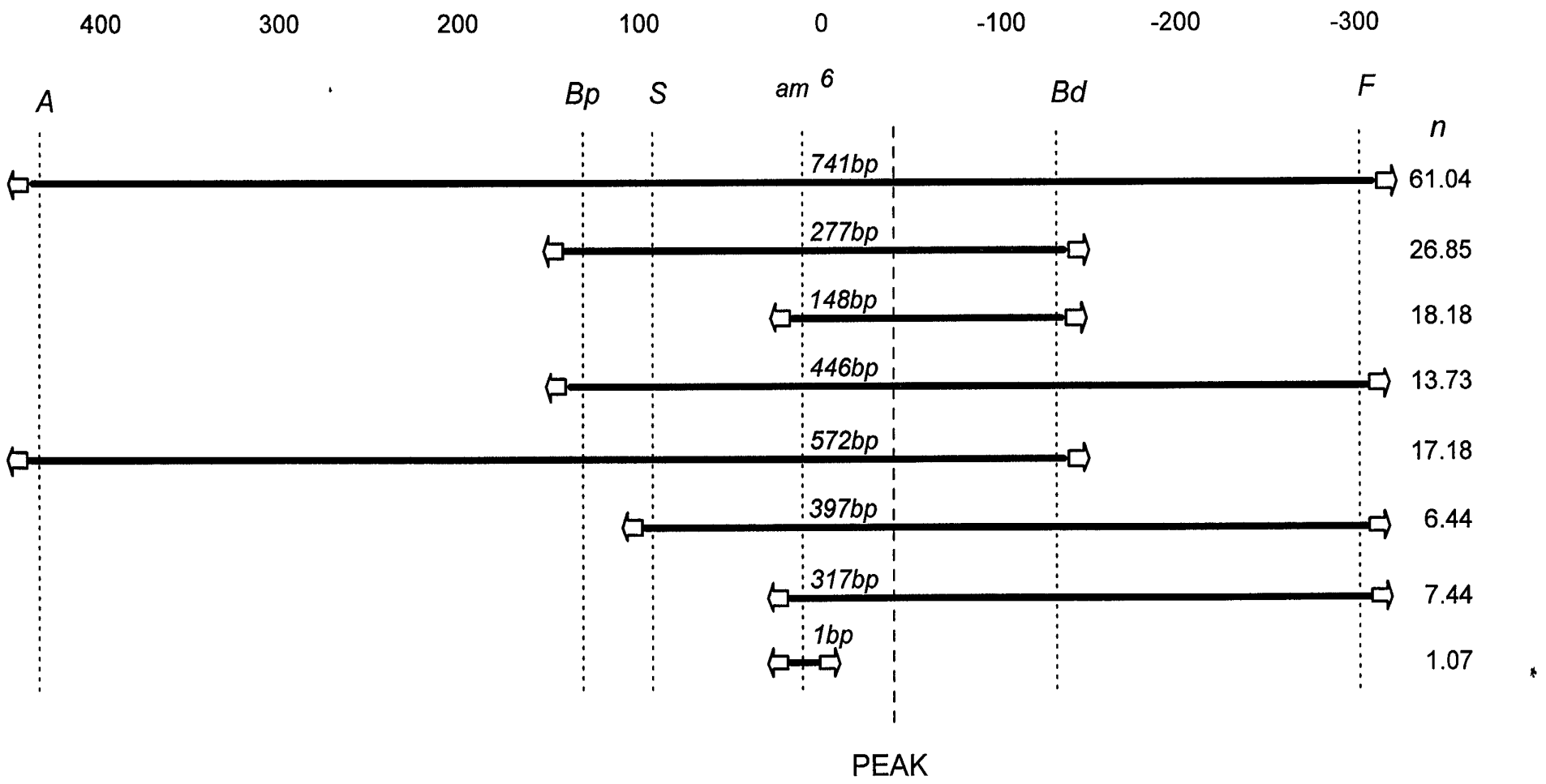
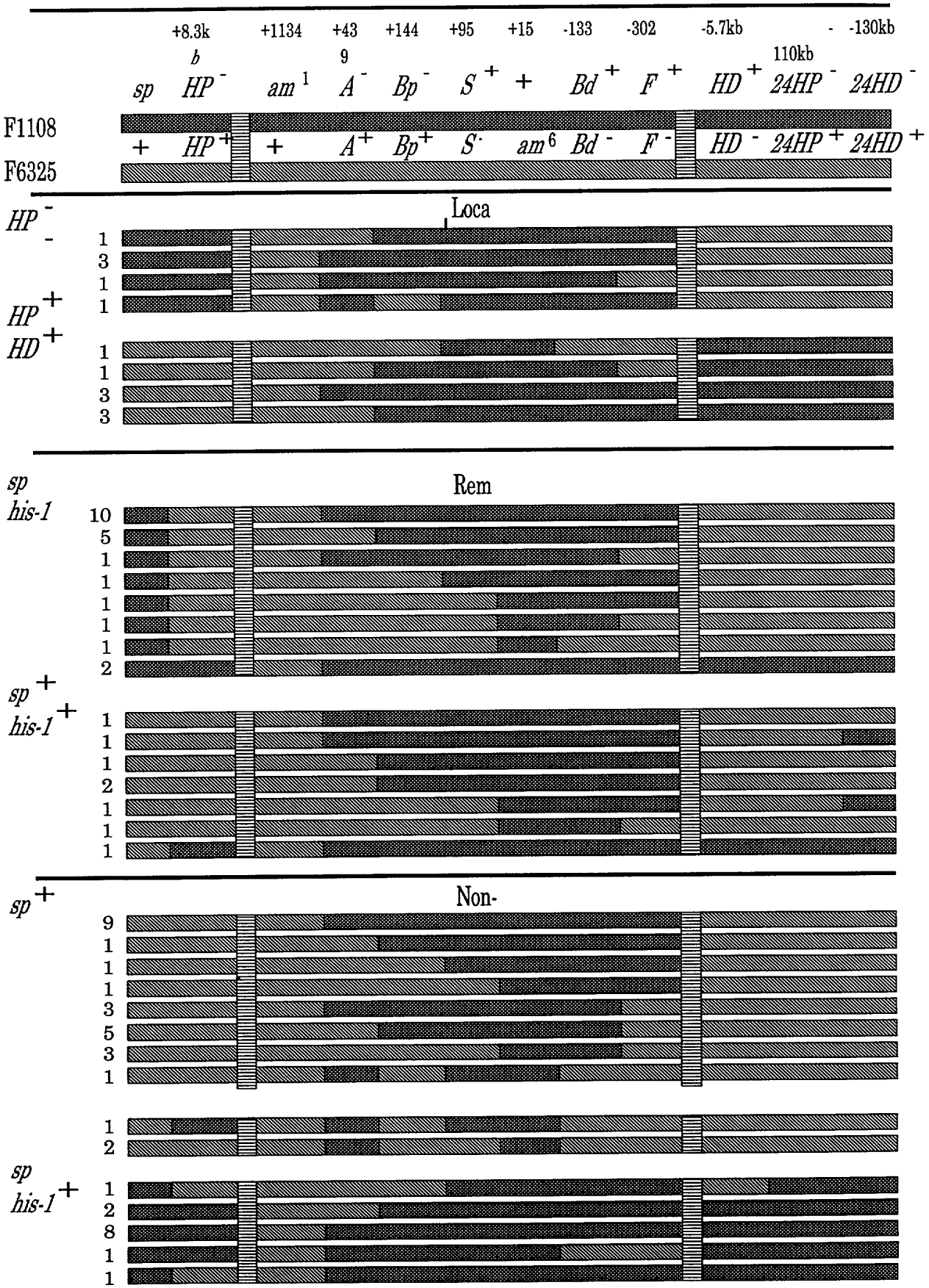

400
300
200
100
0
-100
-200
-300
A
Bp
S
am 6
Bd
F
n
741bp
61.04
277bp
26.85
148bp
18.18
446bp
13.73
572bp
17.18
397bp
6.44
317bp
7.44
1bp
1.07
PEAK

FIG. 14


+8.3k
+1134
+43
+144
+95
+15
-133
-302
-5.7kb
-130kb
b
9
110kb
sp
HP -
am 1
A -
Bp -
S +
+
Bd +
F +
HD +
24HP -
24HD -
F1108
F6325
+ HP +
+ A +
Bp +
S
am 6
Bd -
F -
HD -
24HP +
24HD +
HP -
Loca
HP +
HD +
sp
his-1
Rem
sp +
his-1 +
sp +
Non-
sp
his-1 +